

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/38613 A2

(51) International Patent Classification⁷: C07K 19/00, 14/005, C12N 5/10, 15/33

(21) International Application Number: PCT/EP01/12975

(22) International Filing Date:
9 November 2001 (09.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00124629.7 10 November 2000 (10.11.2000) EP
01109543.7 17 April 2001 (17.04.2001) EP
60/311,876 13 August 2001 (13.08.2001) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- entirely in electronic form (except for this front page) and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(71) Applicant: ARTEMIS PHARMACEUTICALS GMBH [DE/DE]; Neurather Ring 1, 51063 Cologne (DE).

(72) Inventors: KÜHN, Ralf; Hellenthaler Str. 22, 50937 Köln (DE). FELDER, Susanne; Kopenhagener Str. 1, 41469 Neuss (DE). SCHWENK, Frieder; Konradenweg 15, 50999 Köln (DE). KÜTER-LUKS, Birgit; Niederichstr. 10, 50668 Köln (DE). FAUST, Nicole; Dammelsfurter Weg 9, 51063 Rösrath (DE).

(74) Agents: HELBING, Jörg et al.; Von Kreisler Selting Werner, Postfach 10 22 41, 50462 Köln (DE).

(54) Title: MODIFIED RECOMBINASE

(57) Abstract: The present invention concerns a fusion protein comprising a recombinase protein, preferably the site-specific DNA recombinase C31-Int of phage (C31, and a peptide sequence which directs the nuclear uptake of the fusion protein in eucaryotic cells, and the use of this fusion protein to recombine, invert or delete DNA molecules containing recognition sequences for said recombinase in eucaryotic cells at high efficiency. In addition the invention relates to a cell, preferably a mammalian cell which contains recognition sequences for said recombinase in its genome and wherein the genome is recombined by the action of said fusion protein. Moreover, the invention relates to the use of said cell to study the function of genes and for the creation of transgenic organisms to study gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination.



WO 02/38613 A2

Modified Recombinase

5 The present invention concerns a fusion protein comprising a recombinase protein, preferably the site-specific DNA recombinase C31-Int of phage Φ C31, and a peptide sequence which directs the nuclear uptake of the fusion protein in eucaryotic cells, and the use of this fusion protein to recombine, invert or delete DNA molecules containing recognition sequences for said recombinase in
10 eucaryotic cells at high efficiency. In addition the invention relates to a cell, preferably a mammalian cell which contains recognition sequences for said recombinase in its genome and wherein the genome is recombined by the action of said fusion protein. Moreover, the invention relates to the use of said cell to study the function of genes and for the creation of transgenic organisms to study
15 gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination.

20 Background of the invention

The controlled and permanent modification of the genome of eucaryotic cells and organisms is an important method for research applications, e.g. for studying gene function, for medical applications like gene therapy and the creation of disease models and for the design of economically important animals and crops.
25 The basic methods for genome manipulations by the engineering of endogenous genes through gene targeting in murine embryonic stem (ES) cells are well established and used since many years (Capecchi, Trends in Genetics, 5, 70-76 (1989)). Since ES cells can pass mutations induced in vitro to transgenic offspring in vivo it is possible to analyse the consequences of gene disruption in
30 the context of the entire organism. Thus, numerous mouse strains with functionally inactivated genes ("knock-out mice") have been created by this technology and utilised to study the biological function of a variety of genes (Koller et al., Ann. Rev. Immunol., 10, 705 - 730 (1992)). More importantly, mouse mutants created by this procedure (also known as "conventional,
35 complete or classical mutants"), contain the inactivated gene in all cells and

tissues throughout life. Thus, classical mouse mutants represent the best animal model for inherited human diseases as the mutation is introduced into the germline but are not the optimal model to study gene function in adults, e.g. to validate potential drug target genes.

5 A refined method of targeted mutagenesis, referred to as conditional mutagenesis, employs the Cre/loxP site-specific recombination system which enables the temporally and/or spatially restricted inactivation of target genes in cells or mice (Rajewsky et al., J. Clin. Invest., 98, 600 - 603 (1996)). The phage P1 derived Cre recombinase recognises a 34 bp sequence referred to as loxP site
10 which is structured as an inverted repeat of 13 bp separated by an asymmetric 8 bp sequence which defines the direction of the loxP site. If two loxP sites are located on a DNA molecule in the same orientation the intervening DNA sequence is excised by Cre recombinase from the parental molecule as a closed circle leaving one loxP site on each of the reaction products (Kilby et al., TIG, 9, 413-
15 421 (1993)). The creation of conditional mouse mutants initially requires the generation of two mouse strains, one containing two or more Cre recombinase recognition (loxP) sites in its genome while the other harbours a Cre transgene. The former strain is generated by homologous recombination in ES cells as described above, except that the exon(s) of the target gene is (are) flanked by
20 two loxP sites which reside in introns and do not interfere with gene expression. The Cre transgenic strain contains a transgene whose expression is either constitutively active in certain cells and tissues or is inducible by external agents, depending on the promoter region used. Crossing of the loxP-flanked mouse strain with the Cre recombinase expressing strain enables the deletion of the
25 loxP-flanked exons in the genome of doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. Thus, the method allows the analysis of gene function in particular cell types and tissues of otherwise widely expressed genes. Moreover, it enables the analysis of gene function in the adult organism by circumventing embryonic lethality which is often the consequence of
30 complete (germline) gene inactivation. For pharmaceutical research, aiming to validate the utility of genes and their products for drug development, gene inactivation which is inducible in adults provides an excellent genetic tool as this mimicks the biological effects of target inhibition upon drug application.

Since the first description of the concept of conditional gene targeting using the Cre/loxP system in mice in 1994 (Gu et al., Science 265, 103-106 (1994)) this method became increasingly popular among the research community and resulted in a broad collection of genetic tools for biological research in the mouse.

5 More than 30 Cre transgenic mouse strains with various tissue specificities for gene inactivation have been created, including several "deleter" strains which allow to remove the loxP-flanked target gene segment in the male or female germline (Cohen-Tannoudji et al., Mol. Hum. Reprod. 4, 929-938 (1998); Metzger et al., Curr. Op. Biotech., 10, 470-476 (1999)). The need to characterise
10 the expression pattern of Cre mediated recombination in newly generated strains stimulated the construction of a number of "Cre-reporter" strains which harbour a silent reporter gene the expression of which is activated upon Cre-mediated deletion (Nagy, Genesis, 26, 99-109 (2000)). Conditional mouse mutants have been reported for about 20 different genes, many of them could not be studied in
15 adults as their complete inactivation leads to embryonic lethality (Cohen-Tannoudji et al., Mol. Hum. Reprod. 4, 929-938 (1998)).

Great efforts have also been made to control the expression of Cre recombinase in an inducible fashion in mice. After the first demonstration that inducible gene
20 knock-out is feasible in adult mice using an interferon controlled promoter (Kühn et al., Science, 269, 1427-1429 (1995)), mainly two methods were applied to control the activity of Cre recombinase. First, it has been demonstrated that the fusion of Cre with the ligand binding domain of a mutant estrogen receptor allows to control recombinase activity by a specific steroid-like inducer. Several
25 transgenic mouse strains expressing such a fusion protein have been generated and allow to induce gene inactivation in specific tissues (Metzger et al., Curr. Op. Biotech., 10, 470-476 (1999)). Furthermore, the tetracycline-regulated gene expression system has been successfully used to control the expression of Cre in transgenic mice and thus provides a second system for inducible gene
30 inactivation using doxycycline as inducer (Saam et al., J. Biol. Chem. 274, 38071-38082 (1999)).

In addition to the application of Cre/loxP for gene inactivation by deletion of a gene segment this recombination system has been proved to be useful also for a
35 number of other genomic manipulations in ES cells or mice. These include the

conditional activation of transgenes in mice, chromosome engineering to obtain deletion, translocation or inversion, the simple removal of selection marker genes, gene replacement, the targeted insertion of transgenes and the (in)activation of genes by inversion (Nagy, *Genesis*, 26, 99-109 (2000); Cohen-Tannoudji et al., *Mol. Hum. Reprod.* 4, 929-938 (1998)). In conclusion, the Cre/loxP recombination system has been proven to be extremely useful for the analysis of gene function in mice by broadening the methodological spectrum for genome engineering. It can be expected that many of the protocols now established for the mouse may be applied in future also to other animals or plants.

In contrast to the huge diversity of genome manipulations which have been developed for the Cre/loxP system, very limited efforts have been made to develop further site-specific recombination systems for the use in mammalian cells. Alternative recombination systems of different specificity but with an efficiency comparable to Cre/loxP could further enhance the flexibility of genome engineering by the side to side use of two or more systems in the same cell or organism. Furthermore, unidirectional recombination systems which follow a different mechanism than the reversible Cre/loxP-mediated recombination should allow to develop new applications for genome engineering which cannot be performed with the current systems.

The reasons for the almost exclusive use of the Cre/loxP system for site-specific recombination in mammalian cells are readily explained by a number of requirements which must be fulfilled for the efficient use of a recombinase in mammalian cells:

- i) the recombinase should act independent of cofactors like helper proteins,
- ii) it should act independent of the supercoiling status of the target DNA and also on mammalian chromatin,
- iii) it should be efficiently active and stable at a temperature of 37°C, and
- iv) it should recognize a target sequence which is sufficiently long to be unique among large genomes, and it should exhibit a very high affinity to its target site for efficient action (Kilby et al., *TIG*, 9, 413-421 (1993)).

Among the more than 200 described members of the integrase and resolvase/invertase recombinase families only the Cre/loxP system is presently known to fulfill all of these requirements (Nunes-Düby et al., *Nucleic Acids Res.*, 26, 391-406 (1998); Kilby et al., *TIG*, 9, 413-421 (1993); Ringrose et al., *J. Mol. Biol.*, 284, 363 – 384 (1998)). Besides Cre/loxP a few recombinases have been shown to exhibit some activity in mammalian cells but their practical value is presently unclear as their efficiency has not been compared to the Cre/loxP system on the same genomic recombination substrate and in some cases it is known that one or more of the criteria listed above are not met. The best characterised examples are the yeast derived FLP and Kw recombinases which exhibit a temperature optimum at 30°C but which are unstable at 37°C (Buchholz et al., *Nature Biotech.*, 16, 657 – 662 (1998); Ringrose et al., *Eur. J. Biochem.*, 248, 903 – 912). For FLP it has been shown in addition that its affinity to the FRT target site is much lower as compared to the affinity of Cre to loxP sites (Ringrose et al., *J. Mol. Biol.*, 284, 363 – 384 (1998)). Other recombinases which show in principle some activity in mammalian cells are a mutant integrase of phage λ , the integrases of phages Φ C31 and HK022, mutant $\gamma\delta$ -resolvase and β -recombinase (Lorbach et al., *J. Mol. Biol.*, 296, 1175 – 81 (2000); Groth et al., *Proc. Natl. Acad. Sci. USA*, 97, 5995 – 6000 (2000); Kolot et al., *Mol. Biol. Rep.* 26, 207 – 213 (1999); Schwikardi et al., *FEBS Lett.*, 471, 147 – 150 (2000); Diaz et al., *J. Biol. Chem.*, 274, 6634 – 6640 (1999)). Other phage integrase systems include coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase, CisA recombinase, XisF recombinase and transposon Tn4451 TnpX recombinase (Stark et al. *Trends in Genetics* 8, 432-439 (1992); Hatfull & Gridley, in *Genetic Recombination*. Eds. Kucherlipati & Smith, Am. Soc. Microbiol., Washington DC, 357-396 (1988)).

However, the practical value of these recombinases and integrases for use in mammalian cells is limited as their efficiency to recombine mammalian genomic DNA has not been tested or compared with the Cre/loxP system. From the data available it can be assumed that these recombinases are much less effective than the Cre/loxP system.

In a few cases attempts have been made to improve the performance of recombinases in mammalian cells: for FLP a mutant showing improved thermostability and activity at 37°C has been isolated but this mutant is still considerably more heat labile as compared to Cre (Buchholz et al., Nature Biotech., 16, 657 – 662 (1998)). In the case of λ -integrase and $\gamma\delta$ -resolvase the absolute requirement for coproteins and supercoiled DNA could be eliminated by the introduction of specific point mutations (Schwikardi et al. FEBS Lett 471, pp147-50 (2000)).

The import of cytoplasmic proteins into the nucleus of eucaryotic cells through nuclear pores is a regulated, energy dependent process mediated by specific receptors (Görlich et al., Science, 271, 1513 – 1518 (1996)). Proteins which do not possess a signal sequence recognised by the nuclear import machinery are excluded from the nucleus and remain in the cytoplasm. Numerous of such nuclear localisation signal sequences (NLS), which share a high proportion of basic amino acids in common, have been characterised (Boulikas, Crit. Rev. Eucar. Gene Expression, 3, 193 – 227 (1993)), the prototype of which is the 7 amino acid NLS derived from the T-antigen of the SV40 virus (Kalderon et. al, Cell, 39, 499 – 509 (1984)).

It was believed that the fusion of such an NLS peptide to a recombinase possibly would enhance the efficiency of the recombinase by mediating its import into the nucleus and therewith increasing the concentration of the recombinase inside the nucleus. However, for Cre recombinase it has been shown that the addition of the SV-40 T-antigen NLS does not improve its recombination efficiency in mammalian cells (Le et al., Nucleic Acid Res., 27, 4703 – 4709 (1999)). Nevertheless, both Cre and a Cre-NLS-fusion protein are widely used. Schwikardi (Schwikardi et al., FEBS Lett. 471, pp147-50 (2000)) reported a $\gamma\delta$ -resolvase-SV-40 T-antigen NLS fusion protein, which also did not enhance the recombination efficiency.

The level of activity exhibited by recombinases of diverse prokaryotic origin in mammalian cells may be the result of the intrinsic properties of an enzyme depending on parameters like its temperature optimum, its target site affinity, protein structure and stability, the degree of cooperativity, the stability of the

synaptic complex and the dependence on coproteins or supercoiled DNA. Within the specific environment of mammalian cells the activity of a prokaryotic recombinase could be limited by additional factors such as a short half-life of the recombinase transcript, a short half-life of its protein, its inability to act on histone-complexed and higher order structured mammalian genomic DNA, exclusion from the nucleus or the recognition of cryptic splice sites within its mRNA resulting in a nonfunctional transcript. Due to the lack of information on the parameters listed above for almost all recombinases it is presently not possible to rationally optimise their performance in mammalian cells.

Summary of the Invention

The object to be solved by the invention of the present application is the provision of a recombination system alternative to the Cre/loxP system, which has a different specificity but an efficiency comparable to Cre/loxP. Such an alternative recombination system is particularly desirable for all those applications which require more than one potent recombination system for being successfully carried out (e.g. the methods disclosed in PCT/EP01/00060 and PCT/EP00/10162). Most surprisingly, it was found that the above object can be solved by fusing a signal peptide capable directing the nuclear import (hereinafter shortly referred to as nuclear localisation signal sequences (NLS)) to specific recombinases.

In contrast to the wildtype recombinases, the resulting modified recombinases allow a highly efficient recombination of extrachromosomal and chromosomal DNA in mammalian cells, and a highly efficient excision of extrachromosomal and chromosomal DNA-stretches, which are flanked by suitable recognition sites for said modified recombinases.

The present invention thus provides:

(1) A fusion protein (hereinafter also referred to as "modified recombinase") comprising

(a) a recombinase domain comprising a recombinase protein or fragment thereof and

(b) a signal peptide domain being linked to (a) and directing the nuclear import

of said fusion protein in eucaryotic cells,
preferably the activity of the fusion protein in eucaryotic cells is significantly
higher as compared to the activity of the wildtype recombinase corresponding
to the recombinase of the recombinase domain;

- 5 (2) in a preferred embodiment of the fusion protein defined in (1) above, the
recombinase domain comprises an integrase protein, preferably a phage Φ C31
integrase (C31-Int) protein or a mutant thereof;
- (3) a DNA coding for the fusion protein as defined in (1) or (2) above;
- (4) a vector containing the DNA as defined in (3) above;
- 10 (5) a microorganism containing the DNA of (3) above and/or the vector of (4)
above;
- (6) a process for preparing the fusion protein as defined in (1) or (2) above
which comprises culturing a microorganism as defined in (5) above;
- (7) the use of the fusion protein as defined in (1) or (2) above to recombine DNA
15 molecules, which contain recombinase recognition sequences for the
recombinase protein of the recombinase domain, in eucaryotic cells;
- (8) a cell, preferably a mammalian cell containing the DNA sequence of (3) above
in its genome;
- (9) the use of the cell of (8) above for studying the function of genes and for the
20 creation of transgenic organisms;
- (10) a transgenic organism, preferably a transgenic mammal containing the DNA
sequence of (3) above in its genome;
- (11) the use of the transgenic organism of (10) above for studying gene function
at various developmental stages; and
- 25 (12) a method for recombining DNA molecules of cells or organisms containing
recognition sequences for the recombinase protein of the recombinase domain as
defined in (1) or (2) above, which method comprises supplying the cells or
organisms with a fusion protein as defined in (1) or (2) above, or with a DNA
sequence of (3) above and/or a vector of (4) above which are capable of
30 expressing said fusion protein in the cell or organism.

The present invention combines the use of prokaryotic recombinases such as the
C31-Int with a eukaryotic signal sequence which increases its efficiency in
mammalian cells such that it is equal to the widely used Cre/loxP recombination
35 system. The improved recombination system of the present invention provides

an alternative recombination system for use in mammalian cells and organisms which allows to perform the same types of genomic modifications as shown for Cre/loxP, including conditional gene inactivation by recombinase-mediated deletion, the conditional activation of transgenes in mice, chromosome engineering to obtain deletion, translocation or inversion, the simple removal of selection marker genes, gene replacement, the targeted insertion of transgenes and the (in)activation of genes by inversion.

Short Description of Figures

Fig. 1: C31-Int and Cre recombinase expression vectors and a recombinase reporter vector used for transient and stable transfections

Fig. 2: Results of transient transfections of C31 Int and Cre expression vectors and reporter vectors into CHO cells.

Fig. 3: Results of transient transfections of XisA and Ssv recombinase expression vectors with and without nuclear localisation signals and reporter vectors into CHO cells.

Fig. 4: Results of transient transfections of C31-Int and Cre recombinase vectors into a stable reporter cell line.

Fig. 5: In situ detection of β -galactosidase in 3T3(pRK64)-3 cells transfected with recombinase expression vectors

Fig. 6: Test vector for C31-Int mediated deletion, pRK64, and the expected deletion product.

Fig. 7: PCR products generated with the primers P64-1 and P64-4 and sequence comparison.

Fig. 8: ROSA26 locus of the C31 reporter mice carrying a C31 reporter construct.

Fig. 9: *In situ* detection of β -galactosidase in a cryosection of the testis of: (A) a double transgenic mouse carrying both the recombinase and the reporter; and

(B) a transgenic mouse carrying only the reporter as a control.

Detailed Description of the Invention

The "organisms" according to the present invention are multi-cell organisms and
5 can be vertebrates such as mammals (humans and non-human animals including rodents such as mice or rats) or non-mammals (e.g. fish), or can be invertebrates such as insects or worms, or can be plants (higher plants, algi or fungi). Most preferred living organisms are mice and fish.

10 "Cells" and "eucaryotic cells" according to the present invention include cells isolated from the above defined living organism and cultured *in vitro*. These cells can be transformed (immortalized) or untransformed (directly derived from the living organism; primary cell culture).

15 "Microorganism" according to the present invention relates to procaryotes (e.g. *E. coli*) and eucaryotic microorganisms (e.g. yeasts).

According to embodiment (1) of the present invention, the activity of the fusion protein in eucaryotic cells is significantly higher as compared to the activity of
20 the wildtype recombinase corresponding to the recombinase of the recombinase domain. A "significantly higher activity" in accordance with the present invention refers to an increase in activity of at least 50%, preferably at least 75%, more preferably at least 100% relative to the corresponding wildtyp recombinase in eucaryotic cells. A "significantly higher activity" also implies that the resulting
25 fusion protein has at least 25%, preferably at least 50% and more preferably at least 75%, of the activity of Cre/loxP in 3T3 cells with a stably integrated target sequence.

Recombinase proteins which can be used in the recombinase domain of the
30 fusion protein of the present invention (i.e., giving a fusion having a "significantly higher activity" as defined above) include, but are not limited to, a certain type of recombinases belonging to the family of of large serine recombinases (Thorpe et al., Control of directionlty in the site-specific recombination system of the streptomyces phage ϕ C31, Molecular Microbiology 38(2), 232-241 (2000)). This
35 family includes bacteriophage ϕ C31 integrase ("C31-Int"; the amino acid

sequence of said integrase and a DNA sequence coding therefor are shown in SEQ ID NOs:21 and 20, respectively), coliphage P4 recombinase, *Listeria* phage recombinase, bacteriophage R4 Sre recombinase ("R4 Sre" deposited under GI 793758; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:55 and 54, respectively), *Bacillus subtilis* CisA recombinase ("CisA" deposited under GI 142689; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:57 and 56, respectively), XisF recombinase from *Annabaena* sp. Strain PCC 7120 (Cyanobacterium; "XisF" deposited under GI 349678; the amino acid sequence of said integrase and a DNA sequence coding therefor are shown in SEQ ID NOs:59 and 58, respectively), transposon Tn4451 TnpX recombinase ("TnpX" deposited under GI 551135; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:61 and 60, respectively), "XisA" recombinase from *Annabaena* sp. Strain PCC 7120 (Cyanobacterium; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:63 and 62, respectively), "SSV" recombinase from phage of *Sulfolobus shibatae* (the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:65 and 64, respectively); lactococcal bacteriophage TP901-1 recombinase (TP901-1 complete genome deposited under GI 13786531; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:108 and 107, respectively), and the like, or mutants thereof. Other procaryotic recombinases known in the art are also applicable.

A "mutant" of the above recombinases in accordance with the present invention relates to a mutant of the respective original (viz. wild-type) recombinase having a recombinase activity similar (e.g. at least about 90%) to that of said wild-type recombinase. Mutants include truncated forms of the recombinase (such as N- or C-terminal truncated recombinase proteins), deletion-type mutants (where one or more amino acid residues or segments having more than one continuous amino acid residue have been deleted from the primary sequence of the wildtype recombinase), replacement-type mutants (where one or more amino acid residues or segments of the primary sequence of the wildtype recombinase have been replaced with alternative amino acid residues or segments), or combinations thereof.

According to embodiment (2) of the invention, the recombinase domain comprises an integrase protein, preferably a phage Φ C31 integrase (C31-Int) protein or a mutant thereof. Thus, the present invention provides a fusion protein

comprising
 (a) an integrase domain being a C31-Int protein or a mutant thereof, and
 (b) a signal peptide domain being linked to (a) and directing the nuclear import of said fusion protein into eucaryotic cells.

In the fusion protein of embodiment (2), the integrase domain is preferably a C31-Int having the amino acid sequence shown in SEQ ID NO:21 or a C-terminal truncated form thereof. Suitable truncated forms of the C31-Int comprise amino acid residues 306 to 613 of SEQ ID NO:21.

The signal peptide domain (hereinafter also referred to as "NLS") is preferably derived from yeast GAL4, SKI3, L29 or histone H2B proteins, polyoma virus large T protein, VP1 or VP2 capsid protein, SV40 VP1 or VP2 capsid protein, Adenovirus E1a or DBP protein, influenza virus NS1 protein, hepatitis virus core antigen or the mammalian lamin, c-myc, max, c-myb, p53, c-erbA, jun, Tax, steroid receptor or Mx proteins (see Bouliskas, Crit. Rev. Eucar. Gene Expression, 3, 193 - 227 (1993)); simian virus 40 ("SV40") T-antigen (Kalderon et. al, Cell, 39, 499 - 509 (1984)) or other proteins with known nuclear localisation. The NLS is preferably derived from the SV40 T-antigen.

Furthermore, the signal peptide domain preferably has a length of 5 to 74, preferably 7 to 15 amino acid residues. More preferred is that the signal peptide domain comprises a segment of 6 amino acid residues wherein at least 2 amino acid residues, preferably at least 3 amino acid residues are positively charged basic amino acids. Basic amino acids include, but are not limited to, Lysin, Arginin and Histidine. Particularly preferred signal peptides are show in the following table.

Organism	Sequence/(SEQ ID NO:)
yeast GAL4	MKx11CRLKKLKCSKEKPKCAKCLKx5Rx3KTKR (24)
yeast SKI3	IKYFKKFPKD (25)

	yeast L29	MTGSKTRKHRGSGA	(26)
		(MTGSKHRKHPGSGA)	(27)
	yeast histone H2B	(GKKRСКА)	(28)
	polyoma virus large T protein	(PKKAREDVSRKRPR)	(29)
5	polyoma virus VP1 capsid protein	(APKRKSGVSKC)	(30)
	polyoma virus VP2 capsid protein	(EEDGPQKKRRL)	(31)
	SV40 VP1 capsid protein	(APTKRKGS)	(32)
	SV40 VP2 capsid protein,	(PNKKKРK)	(33)
	Adenovirus E1a protein	(KRPRP)	(34)
10		(CGGLSSKRPRP)	(35)
	Adenovirus DBP protein	(PPKKRMRRRIEPKKKKKRP)	(36)
	influenza virus NS1 protein	(PFLDRLRRDQK)	(37)
		(PKQKRКMAR)	(38)
	human laminaA	(SVTKKРKLE)	(39)
15	human c-myc	(CGGAAKRVKLD)	(40)
		(PAAKRVKLD)	(41)
		(RQRRNELKRSP)	(42)
	HUMAN max	(PQSRKKLR)	(43)
	HUMAN c-myb	(PLLKKIKQ)	(44)
20	HUMAN p53	(PQPKKKP)	(45)
	HUMAN c-erba	(SKRVAKRKL)	(46)
	VIRAL jun	(ASKSRKРKL)	(47)
	HUMAN Tax	(GGLCSARLHRHALLAT)	(48)
	Mammalian glucocorticoid receptor	(RKTKKKIK)	(49)
25	HUMAN ANDROGEN RECEPTOR	(RKLKKLGN)	(50)
	MAMMALIAN ESTROGEN RECEPTOR	(RKDRRGGR)	(51)
	Mx proteins	(DTREKKKFLKRRLLRLDE)	(52)
	SV40 T-antigen	(PKKKRKV)	(53)

30 The most preferred signal peptide domain is that of SV40 T-antigen having the sequence Pro-Lys-Lys-Lys-Arg-Lys-Val.

The signal peptide domain may be linked to the N-terminal or C-terminal of the integrase domain or may be integrated into the integrase domain, preferably the
 35 signal peptide domain is linked to the C-terminal of the integrase domain. With

regard to phage Φ C31 integrase protein of embodiment (2) of the invention it was found that the fusion of an NLS-peptide to the C-terminus of the integrase provided a much higher increase of activity as compared to the fusion of the same NLS-peptide to the N-terminus of the integrase (see Example 1, figures 3 and 4).

According to the present invention, the signal peptide domain may be linked to the integrase domain directly or through a linker peptide. Suitable linkers include peptides having from 1 to 30, preferably 1 to 15 amino acid residues, said amino acid residues being essentially neutral amino acids such as Gly, Ala and Val.

The most preferred fusion protein of the present invention comprises the amino acid sequence shown in SEQ ID NO:23 (a suitable DNA sequence coding for said fusion protein being shown in SEQ ID NO:22).

Further preferred fusion proteins of the present invention are "NLS-XisA" and "NLS-SSV" (having the NLS-peptide fused to the N-terminus of the recombinases) as shown in SEQ ID NO:67 and 69, respectively (suitable DNA sequences coding for said fusion proteins being shown in SEQ ID NO:66 and 68, respectively).

In embodiments (7), (8), (10) and (12) of the invention the DNA molecules, the cell or transgenic organism may also contain recognition sequences for the recombinase protein of the recombinase domain. Thus, when utilizing the fusion protein of embodiment (2), the C31-Int recognition sequences attP and attB are present in DNA molecules, the cell or transgenic organism.

The term "mammal" as used in embodiment (10) of the invention includes non-human mammals (viz. animals as defined above) and humans (if such subject matter is patentable with the respective patent authority).

Since the modified recombinase of the invention, in particular the modified C31-Int, acts in mammalian cells as efficient (or at least almost as efficient) as the widely used Cre/loxP system it can be used for a large variety of genomic modifications (including the methods disclosed in PCT/EP01/00060 and

PCT/EP00/10162, the content of which is herewith incorporated by reference). Concerning embodiment (11) it is to be noted that the mammals of embodiment (10) can be used to study the function of genes, e.g. in mice, by conditional gene targeting. For this purpose suitable recognition sequences - when utilizing the
5 fusion protein of embodiment (2), one attP and one attB site (C31-Int recognition sequences) in the same orientation - can be introduced into introns of a gene by homologous recombination of a gene targeting vector in ES cells such that the two sites flank one or more exons of the gene to be studied but do not interfere with gene expression. A selection marker gene, needed to isolate recombinant ES
10 cell clones, can be flanked by two recognition sites of another recombinase such as loxP or FRT sites to enable deletion of the marker gene upon transient expression of the respective recombinase in ES cells. These ES cells can be used to generate *germline chimaeric mice* which transmit the target gene modified by att sites to their offspring and allow to establish a modified mouse strain. The
15 crossing of this strain with a C31-Int recombinase transgenic line or the application of C31-Int protein will result in the deletion of the att-flanked gene segment from the genome of doubly transgenic offspring and the inactivation of the target gene in doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. The C31-Int transgenic strain contains a transgene
20 whose expression is either constitutively active in certain cells and tissues or is inducible by external agents, depending on the promoter region used. If an attB and an attP site are placed into the genome in opposite orientation C31-Int mediated recombination results in the irreversible inversion of the flanked gene segment leading the functional loss of one or more exons of the target gene.
25 Thus, the method allows the analysis of gene function in particular cell types and tissues of otherwise widely expressed genes and circumvents embryonic lethality which is often the consequence of complete (germline) gene inactivation. For the validation of genes and their products for drug development, gene inactivation which is inducible in adults provides an excellent genetic tool as this mimicks the
30 biological effects of target inhibition upon drug application. If a pair of attB/P sites is placed in the same or opposite orientation into a chromosome at large distance using two gene targeting vectors, C31-Int recombination allows to delete or invert chromosome segments containing one or more genes, or chromosomal translocations if the two sites are located on different
35 chromosomes. In another application of the method a pair of attB/P sites is

placed in the same orientation within a transgene such that the deletion of the att-flanked DNA segment results in gene expression, e.g. of a toxin or reporter gene for cell lineage studies, or in the inactivation of the transgene.

5 In addition, according with embodiment (12) of the invention, the recombination system of embodiment (1), in particular the C31-Int recombination system of embodiment (2), can also be used for the site specific integration of foreign DNA into the genome of mammalian cells, e.g. for gene therapy. For this purpose, and if the C31-Int recombination system of embodiment (2) is utilized, only one attB
10 (or attP) site is initially introduced into the genome by homologous recombination, or an endogenous genomic sequence which resembles attB or attP is used. The application of a vector containing an attP (or attB) site to such cells or mice in conjunction with the expression of C31-Int recombinase will lead to the site specific integration of the vector into the genomic att site.

15

Thus, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination. Said process possesses the following advantages over current technology:

- 20 (i) the modified recombinase, in particular the modified C31-Integrase, allows to recombine extrachromosomal and genomic DNA in mammalian cells at much higher efficiency as compared to the use of its wildtype form;
- (ii) the modified recombinase, in particular the modified C31-Integrase, is the
25 first described alternative recombination system with equal efficiency to Cre/loxP for the recombination of chromosomal DNA in mammalian cells.

The appended figures further explain the present invention:

30 Figure 1 shows C31-Int and Cre recombinase expression vectors and a recombinase reporter vector used for transient and stable transfections.

A-D: Mammalian expression vectors for recombinases which contain the CMV immediate early promoter followed by a hybrid Intron, the coding region of the recombinase to be tested, and an artificial polyadenylation signal sequence (pA).

A: pCMV-C31Int(wt) containing the nonmodified (wildtype) 1.85 kb coding region of C31-Int as found in the genome of phage Φ X31.

B: pCMV-C31Int(NNLS) containing a modified C31-Int gene coding for the full length C31-Int protein with a N-terminal fusion to the SV40 virus large T antigen nuclear localisation signal (NLS).

C: pCMV-C31Int(CNLS) containing a modified C31-Int gene coding for the full length C31-Int protein with a C-terminal fusion to the SV40 virus large T antigen nuclear localisation signal (NLS).

D: pCMV-Cre contains the 1.1 kb Cre coding region with an N-terminal fusion to the SV40 T antigen NLS.

E: Recombination substrate vector pRK64 contains a SV40 promoter region followed by a 1.1 kb cassette consisting of the coding region of the puromycin resistance gene and a polyadenylation signal sequence, flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site of C31-Int. pRK64 contains in addition two Cre recognition (loxP) sites in direct orientation next to the att sites.

Figure 2 shows results of transient transfections of C31-Int and Cre recombinase and reporter vectors into CHO cells.

All transfections were performed with a fixed amount of the reporter plasmid pRK64 and 0.5 ng or 1 ng of the recombinase expression plasmids pCMV-C31-Int(wt) (samples 4-5), pCMV-C31-Int(NNLS) (samples 6-7), pCMV-C31-Int(CNLS) (samples 8-9) or pCMV-Cre (samples 10-11). Negative controls: transfection with pRK64 (sample 3) or pUC19 alone (sample 1). Positive control: transfection with the Cre-recombined reporter pRK64(Δ Cre) (sample 2).

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)), the RLU from the assay for Luciferase, the ratio of the β -galactosidase and Luciferase values with standard deviation (RLU $\times 10^5$ (Gal/Luc)), and the relative activity of the various recombinases as compared to the positive control defined as 1.

Figure 3 shows results of transient transfections of XisA and Ssv recombinases and reporter vectors into CHO cells.

All transfections were performed with fixed amounts of the reporter plasmids pPGKnif (for XisA) and pPGKattA (for SSV) and 25 ng or 100 ng of the

recombinase expression plasmids pCMV-XisA, pCMV-XisA(NNLS) and 10 ng or 20 ng of the expression plasmids pCMV-Ssv and pCMV-Ssv(NNLS). Negative controls: transfection with pPGKnif or pPGKattA alone.

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)), the RLU from the assay for Luciferase, the ratio of the β -galactosidase and "Luciferase" values with standard deviation (RLU $\times 10^5$ (Gal/Luc)).

Figure 4 shows results of transient transfections of recombinase vectors into a stable reporter cell line.

All transfections were performed with a NIH 3T3 derived clone containing stably integrated copies of the pRK64 recombination substrate vector. Either 32 ng or 64 ng of the recombinase expression plasmids pCMV-C31-Int(wt) (samples 2-3), pCMV-C31-Int(NNLS) (samples 4-5), pCMV-C31-Int(CNLS) (samples 6-7) or pCMV-Cre(NNLS) (samples 8-9). Negative control: transfection with pUC19 alone (sample 1).

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)) and the relative activity of the various recombinases as compared to the value obtained with pCMV-Cre(NNLS) defined as 1.

Figure 5 shows the in situ detection of β -galactosidase in 3T3(pRK64)-3 cells transfected with recombinase expression vectors.

The Cre and C31-Int recombinase reporter cell line 3T3(pRK64)-3 was either not transfected with DNA (A), transfected with the Cre expression vector pCMV-Cre (B) or with the C31-Int expression vector pCMV-C31-Int(CNLS). Two days after tranfection the cells were fixed and incubated with the histochemical X-Gal assay which develops a blue stain in β -galactosidase positive cells indicating recombinase mediated activation of the reporter gene.

Figure 6 shows the test vector for C31-Int mediated deletion, pRK64, and the expected product of deletion, pRK64(Δ Int).

Plasmid pRK64 contains the 1.1 kb cassette of the coding region of the puromycin resistance gene and a polyadenylation signal, which is flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site (large triangles) of C31-

Int. These attB and attP sites are oriented in the same way to each other (thick black arrows) which is used by the Φ X31 phage to integrate into the bacterial genome. In addition, the cassette is flanked by two Cre recombinase recognition (loxP) sites in the same orientation (black small triangles). For better orientation the half sites of the att sequences are labelled by a direction (thin arrow) and numbered 1-4. The 3 bp sequence within the att sites at which recombination occurs is framed by a box. The positions at which the PCR primers P64-1 and P64-4 hybridise to the pRK64 vector are indicated by arrows, pointing into the 3' direction of both oligonucleotides.

PRK64(Δ Int) depicts the deletion product expected from the C31-Int mediated recombination between the att sites of pRK64. The recombination between a pair of attB/attP sites generates an attR site remaining on the parental DNA molecule while the puromycin cassette is excised. In this configuration the primers P64-1 and P64-4 will amplify a PCR product of 630 bp from pRK64(Δ Int).

Figure 7 shows PCR products generated with the primers P64-1 and P64-4 and a sequence comparison of the PCR product.

A: Analysis of PCR products on an agarose gel from PCR reactions using the Primers P64-1 and P64-4 on DNA extracted from MEF5-5 cells transfected 2 days before with plasmid pRK64 alone (lane 4), with pRK64 + CMV-Cre (lane 3), with pRK64 + pCMV-C31-Int(wt) (lane 2), and from a control reaction which did not contain cellular DNA (lane 1). The product with an apparent size around 650 bp, as compared to the size marker used, from lane 2 was excised from the agarose gel and purified. The PCR product was cloned into a sequencing plasmid vector and gave rise to the plasmid pRK80d. The insert of this plasmid was sequenced using reverse primer (seq80d) and compared to the predicted sequence of the pRK64 vector after C31-Int mediated deletion of the att flanked cassette, pRK64(Δ Int). The cloned PCR product shows a 100% identity with the predicted attR sequence after deletion. The generated attR site is shown in a box, with the same sequence designation used in Figure 5. The nucleotide positions (pos.) of the compared sequences pRK64(Δ Int) and Seq80d are indicated.

Figure 8 shows the modified ROSA26 locus of C31 reporter mice (Seq ID NO:106). A recombination substrate has been inserted in the ROSA26 locus. The substrate consists of a splice acceptor (SA) followed by a cassette consisting of

the hygromycin resistance gene driven by a PGK promoter and flanked by the recombination sites attB and attP. In addition the reporter contains two Cre recognition sites (loxP) in direct orientation next to the att sites. This cassette is followed by the coding region for β -galactosidase, which is only expressed when the hygromycin resistance gene has been deleted by recombination.

Figure 9 shows the in situ detection of β -galactosidase activity. A cryosection of the testis of a double transgenic mouse carrying both the C31-int recombinase and the recombination substrate was stained with X-Gal (A). The blue colour indicates recombination of the substrate, which leads to the expression of β -galactosidase. As a control a cryosection of testis of a transgenic mouse carrying only the recombination substrate was stained with X-Gal (B).

The present invention is further illustrated by the following Examples which are, however, not to be construed as to limit the invention.

Examples

Example 1

As compared to Cre recombinase the wildtype form of C31-Int exhibits a significantly lower recombination activity in mammalian cells which falls in the range of 10 – 40% of Cre, depending on the assay system used (see below). As a measure which may increase C31-Int efficiency in eukaryotic cells we designed mammalian expression vectors for N- or C-terminal fusion proteins of C31-Int with a peptide was designed which is recognised by the nuclear import machinery. The recombination efficiency obtained by this modified C31-Int recombinase in mammalian cells was compared side by side to the unmodified (wildtype) form of C31-Int and to Cre recombinase. For the quantification of recombinase activities the expression vectors were transiently introduced into a mammalian cell line together with a reporter vector which contains C31-Int and Cre target sites and leads to the expression of β -galactosidase upon recombinase mediated deletion of a vector segment flanked by recombinase recognition sites.

A. Plasmid constructions:

Construction of the recombination test vectors pPGKnif and pPGKattA: first a nifD site (Haselkorn, Annu Rev.Genet. 26, 113-130 (1992)) generated by the annealing of the two synthetic oligonucleotides nifD3 (SEQ ID NO:89) and nifD4 (SEQ ID NO:90), was ligated into the BamHI restriction site of the vector PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 3' of its puromycin resistance gene and loxP site, giving rise to plasmid pPGKnifD3' (SEQ ID NO:79). Next, another nifD site, generated by the annealing of the two synthetic oligonucleotides nifD1 (SEQ ID NO:87) and nifD2 (SEQ ID NO:88), was ligated into the BstBI restriction site of plasmid pPGKnifD3', upstream of the puromycin resistance gene and loxP site, giving rise to plasmid pPGKnifD (SEQ ID NO:78). For pPGKattA (Muskhelishvili et al., Mol.Gen.Genet. 237, 334-342 (1993)) first a 352bp-fragment was amplified from genomic DNA from the thermophilic bacterium Sulfolobus shibatae (DSM-5389, DSMZ Braunschweig-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) with oligonucleotides SSV5 (SEQ ID NO:96) and SSV6 (SEQ ID NO:97) including restriction sites for BamHI and BstBI. The amplified fragment was cloned into the BamHI site of the vector PSV-Pax1 giving rise to plasmid pPGKattA1 (SEQ ID NO:82), subsequently the same 352 bp-fragment was cloned into the BstBI site of pPGKattA1 giving rise to the plasmid pPGKattA2 (SEQ ID NO:83). The sequence and orientation of both nifD sites and attA sites was confirmed by DNA sequence analysis. In pPGKnifD/pPGKattA2 the newly cloned nifD/attA sites (positions 535-619 and 1722-1787/ positions 6718-7081 and 12-363) are in the same orientation flanking the puromycin resistance gene and the SV40 early polyadenylation sequence. The nifD/attA sites are followed by loxP sites in the same orientation (positions 623 - 656 and 1794 - 1827/ positions 7085-7118 and 369-402). The puromycin cassette is transcribed from the SV40 early enhancer/promoter region and followed by the coding region for E. coli β -galactosidase and the SV40 late region polyadenylation sequence.

Construction of XisA and SSV expression vectors: First the XisA gene of cyanobacterium PCC7120 was amplified by PCR from genomic DNA from Nostoc strain PCC7120 (CNCM-Collection Nationale de Cultures de Microorganismes, Institut Pasteur, Paris) using the primers XisA1 (SEQ ID NO:84) and XisA3 (SEQ ID NO:86), and XisA1 (SEQ ID NO:84) and XisA2 (SEQ ID NO:85) (with NLS).

The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmids pRK42a and pRK43 (with NNLS). The DNA sequence of the insert was determined and found to be identical to the published XisA sequence (Genbank GI:3953452) apart from four silent point mutations. The XisA gene was isolated as a 1.4 kb fragment from pRK42a and pRK43 by digestion with NotI and ligated into the generic mammalian expression vector pRK50 (see below), opened with NotI, giving rise to the XisA expression vectors pCMV-XisA (SEQ ID NO:76) and pCMV-XisA(NNLS) (SEQ ID NO:77). pCMV-XisA(wt) contains a Cytomegalovirus immediated early gene promoter (position 1 – 616), a 240 bp hybrid intron (position 716 – 953), the XisA gene (position 974 – 2392), and a synthetic polyadenylation sequence (position 2413 – 2591).

The SSV gene was amplified from genomic DNA from the thermophilic bacterium *Sulfolobus shibatae* (DSM-5389, DSMZ Braunschweig- Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) in two PCR steps because of an internal attP sequence.

First, two overlapping PCR fragments were created with the oligonucleotides SSV1-1 (SEQ ID NO:91) (or SSV1-2 for the SSV(NNLS) gene) and SSV2 (SEQ ID NO:93) and oligonucleotides SSV3 (SEQ ID NO:94) and SSV4 (SEQ ID NO:95).

Using these overlapping fragments as template, a 1000bp fragment containing the complete SSV coding sequence was amplified with primers SSV1-1 (or SSV1-2 for the SSV(NNLS) gene) and SSV4. The 5' 620 bp-fragments of these PCR products were isolated by digestion with NotI-XhoI and cloned into vector pBluescript II KS giving rise to plasmids pRK47 and pRK48 (with NLS). The 3' 380 bp fragment generated by XhoI-digestion was cloned into the XhoI restriction site of vector pBluescript II KS giving rise to the plasmid pBS-SSVs (SEQ ID NO:72). The 380bp SSV-fragment was then isolated by digestion of pBS-SSVs with XhoI and ligated into pRK47 and pRK48 opened by XhoI giving rise to plasmids pBS-SSV3 (SEQ ID NO:70) and pBS-SSV4 (SEQ ID NO:71) (with NLS) containing the complete SSV gene. Sequencing of the plasmids confirmed one point mutation in both plasmids. Therefore 312 bp/ 91 bp fragments generated by digestion with EcoRV-SmaI/ EcoRV-XhoI of another clone of pRK47 were exchanged in plasmids pBS-SSV3/ pBS-SSV4. Sequences were confirmed by sequencing. The SSV gene was isolated from pRK47 and pRK48 by digestion with NotI and KpnI and ligated into the generic mammalian expression vector

pRK50 (see below), opened with NotI and SalI, giving rise to the SSV expression vectors pCMV-SSV(wt) (SEQ ID NO:74) and pCMV-SSV(NNLS) (SEQ ID NO:75).

Construction of the recombination test vector pRK64: first an attB site (Thorpe et al. Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-4 (SEQ ID NO:1) and C31-5 (SEQ ID NO:2), was ligated into the BstBI restriction site of the vector PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 5' of its puromycin resistance gene and loxP site, giving rise to plasmid pRK52. The sequence and orientation of the cloned attB site was confirmed by DNA sequence analysis. Next, an attP site (Thorpe et al. Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-6 (SEQ ID NO:3) and C31-7-2 (SEQ ID NO:4), was ligated into the BamHI restriction site of plasmid pRK52, downstream of the puromycin resistance gene and loxP site, giving rise to plasmid pRK64 (SEQ ID NO:5). The sequence and orientation of the attP site was confirmed by DNA sequence analysis. In pRK64 the newly cloned attB (position 348 - 431) and attP (position 1534 - 1617) sites are in the same orientation flanking the puromycin resistance gene and the SV40 early polyadenylation sequence. The attB and attP sites are followed by loxP sites in the same orientation (positions 435 - 469 and 1624 - 1658). The puromycin cassette is transcribed from the SV40 early enhancer/promoter region and followed by the coding region for E. coli β -galactosidase and the SV40 late region polyadenylation sequence.

Construction of C31-Int expression vectors: First the C31-Int gene of phage Φ C31 was amplified by PCR from phage DNA (DSM-49156, DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) using the primers C31-1 (SEQ ID NO:6) and C31-3 (SEQ ID NO:7). The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmid pRK40. The DNA sequence of the 1.85 kb insert was determined and found to be identical to the published C31-Int gene (Kuhstoss et al., J. Mol. Biol. 222, 897-908 (1991)), except for an error in the stop codon. This error was repaired by PCR amplification of a 300 bp fragment from plasmid pRK40 using the primers C31-8 (SEQ ID NO:8) and C31-9 (SEQ ID NO:9), which provide a

corrected Stop codon. The ends of this PCR fragment were digested with Eco47III and XhoI, the fragment was ligated into plasmid pRK40 and opened with Eco47III and XhoI to remove the fragment containing the defective stop codon. The resulting plasmid pRK55 contains the correct C31-Int gene as confirmed by DNA sequence analysis.

The C31-Int gene was isolated from pRK55 as 1.85 kb fragment by digestion with NotI and XhoI and ligated into the generic mammalian expression vector pRK50 (see below), opened with NotI and XhoI, giving rise to the C31-Int expression vector pCMV-C31-Int(wt). pCMV-C31-Int(wt) (SEQ ID NO:10) contains a 700 bp cytomegalovirus immediated early gene promoter (position 1 – 700), a 270 bp hybrid intron (position 701 – 970), the C31-Int gene (position 978 – 2819), and a 189 bp synthetic polyadenylation sequence (position 2831 – 3020).

For the construction of pCMV-C31-Int(NNLS) a 1.5 kb fragment was amplified by PCR from phage DNA using oligonucleotides C31-2 (SEQ ID NO:98) and C31-3 (SEQ ID NO:7). The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmid pRK41 (SEQ ID NO: 99). A 1100 bp fragment generated by digestion of plasmid pRK41 with NcoI and NotI was then ligated into plasmid pRK55 (SEQ ID NO:80), opened with NcoI and NotI, giving rise to the plasmid pRK63 (SEQ ID NO:81). The C31-Int gene with N-terminal NLS was isolated as a 1.8 kb fragment from pRK63 by digestion with NotI and XhoI and ligated into the mammalian expression vector pRK50, opened with NotI and XhoI, giving rise to the C31-Int expression vector pCMV-C31-Int(NNLS). pCMV-C31-Int(NNLS) (SEQ ID NO:73) contains a 700 bp Cytomegalovirus immediated early gene promoter (position 1 – 700), a 270 bp hybrid intron (position 701 – 970), the C31-Int gene with N-terminal NLS (position 976 – 2838), and a 189 bp synthetic polyadenylation sequence (position 2851 – 3040).

For the construction of pCMV-C31-Int(CNLS), the 3'-end of the C31-Int gene was amplified from pCMV-C31-Int(wt) as a 300 bp PCR fragment using the primers C31-8 (SEQ ID NO:8) and C31-2-2 (SEQ ID NO:11). Primer C31-2-2 modifies the 3'-end of the wildtype C31-Int gene such that the stop codon is replaced by a sequence of 21 basepairs coding for the SV40 T-antigen nuclear localisation sequence of 7 amino acids (Prolin-Lysin-Lysin-Lysin-Arginin-Lysin-Valin) (Kalderon et. al, Cell, 39, 499 – 509 (1984)), followed by a new stop

codon. The ends of this 300 bp PCR fragment were digested with with *Eco47III* and *XhoI*, the fragment was ligated into plasmid pCMV-C31-Int(wt) and opened with *Eco47III* and *XhoI* to replace the 3'-end of the wildtype C31-Int gene resulting in the plasmid pCMV-C31-Int(CNLS). The identity of the new gene
5 segment was verified by DNA sequence analysis. pCMV-C31-Int(CNLS) (SEQ ID NO:12) contains a 700 bp cytomegalovirus immediated early gene promoter (position 12 – 711), a 270 bp hybrid intron (position 712 – 981), the modified C31-Int gene (position 989 – 2851), and a 189 bp synthetic polyadenylation sequence (position 2854 – 3043).

10 To generate the Cre expression plasmid pCMV-Cre (SEQ ID NO:13), the coding sequence of Cre recombinase (Sternberg et al., J. Mol. Biol., 187, 197 – 212 (1986)) with a N-terminal fusion of the 7 amino acid SV40 T-antigen NLS (see above) was recovered from plasmid pgk-Cre and cloned into the *NotI* and *XhoI*
15 sites of plasmid pRK50. PRK50 (SEQ ID NO:14) is a generic expression vector for mammalian cells based on the cloning vector pNEB193 (New England Biolabs Inc, Beverly, MA, USA). PRK50 was built by insertion into pNEB193 of a 700 bp cytomegalovirus immediated early gene (CMV-IE) promoter (position 1–700) from plasmid pIRESHyg (GenBank#U89672; Clontech Laboratories Inc, Palo Alto,
20 CA, USA), a synthetic 270 bp hybrid intron (position 701-970), consisting of a adenovirus derived splice donor and an IgG derived splice acceptor sequence (Choi et al., Mol. Cell. Biol., 11, 3070 – 3074 (1991)), and a 189 bp synthetic polyadenylation sequence (position 1000-1188) build from the polyadenylation consensus sequence and 4 MAZ polymerase pause sites (Levitt et al.,
25 Genes&Dev., 3, 1019 – 1025 (1989); The EMBO J. 13, 5656 – 5667 (1994)). The positive control plasmid pRK64(Δ Cre) (SEQ ID NO:15) was generated from pRK64 by transformation into the Cre expressing E. coli strain 294-Cre (Buchholz et al., Nucleic Acids Res., 24, 3118 – 3119 (1996)).

30 One of the transformed subclones was confirmed for the Cre mediated deletion of the loxP-flanked cassette by restriction mapping and further expanded. Plasmid pUC19 is a cloning vector without eukaryotic control elements used to equalise DNA amounts for transfections (GenBank#X02514; New England Biolabs Inc, Beverly, MA, USA). All plasmids were propagated in DH5 α E. coli cells (Life
35 Technologies GmbH, Karlsruhe, Germany) grown in Luria-Bertani medium and

purified with the plasmid DNA purification reagents "Plasmid-Maxi-Kit" (Quiagen GmbH, Hilden, Germany) or "Concert high purity plasmid purification system" (Life Technologies GmbH, Karlsruhe, Germany). Following purification, the plasmid DNA concentrations were determined by absorption at 260 nm and 280 nm in UVette cuvettes (Eppendorf-Netheler-Hinz GmbH, Hamburg, Germany) using a BioPhotometer (Eppendorf-Netheler-Hinz GmbH, Hamburg, Germany) and the plasmids were diluted to the same concentration; finally these were confirmed by separation of 10 ng of each plasmid on an ethidiumbromide-stained agarose gel.

B. Cell culture and transfections: Chinese hamster ovary (CHO) cells (Puck et al., J. Exp. Med., 108, 945 (1958)) were obtained from the Institute for Genetics (University of Cologne, Germany) as a population adapted to growth in DMEM medium. The cells were grown in DMEM/Glutamax medium (Life Technologies) supplemented with 10% fetal calf serum at 37°C, 10% CO₂ in humid atmosphere and passaged upon trypsinisation. One day before transfection 10⁶ cells were plated into a 48-well plate (Falcon). For the transient transfection of cells with plasmids each well received into 250 µl of medium a total amount of 300 ng supercoiled plasmid DNA complexed before with the FuGene6 transfection reagent (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturers protocol. Each 300 ng DNA preparation (Fig.2 sample 4 to 11) contained 50 ng of the luciferase expression vector pUHC13-1 (Gossen et al., Proc Natl Acad Sci USA., 89 5547-5551 (1992)), 50 ng of the substrate vector pRK64, 0.5 ng or 1 ng of one of the recombinase expression vectors pCMV-C31Int(wt), pCMV-C31Int(NNLS), pCMV-C31Int(CNLS) or pCMV-Cre and 199 ng or 199.5 ng of pUC19 plasmid, except for the controls which received 50 ng of pUHC13-1 together with 50 ng of pRK64 (sample 3) or pRK64(Δcre) (sample 2) and 200 ng pUC19, or 50 ng pUHC13-1 with 250 ng pUC19 (sample 1). Transfections of Ssv and XisA recombinases (Fig.3) also contained 50 ng of the luciferase expression vector pUHC13-1, 50 ng of substrate vectors pPGKattA and pPGKnif and 10 ng or 20 ng of recombinase expression vector pCMV-SSV or pCMV-SSV(NNLS) or 25 ng or 100 ng of expression vectors pCMV-XisA/ pCMV-XisA(NNLS). Plasmid pUC19 was added to a total amount of 300 ng plasmid DNA. As the C31-Int expression vectors are 15% larger in size than pCMV-Cre and the same amounts of DNA of the three plasmids were used for transfection, the

samples with C31-Int vectors received 15% less plasmid molecules as compared to the samples with Cre expression vector. The β -galactosidase values from C31-Int transfected samples by this value were not corrected and thus is a slight underestimation of the calculated C31-Int activities. For each sample to be tested four individual wells were transfected. One day after the addition of the DNA preparations each well received additional 250 ml of growth medium. The cells of each well were lysed 48 hours after transfection with 100 ml lysate reagent supplemented with protease inhibitors (Roche Diagnostics). The lysates were centrifuged and 20 ml were used to determine the β -galactosidase activities using the β -galactosidase reporter gene assay (Roche Diagnostics) according to the manufacturers protocol in a Lumat LB 9507 luminometer (Berthold). To measure luciferase activity, 20ml lysate was diluted into 250ml assay buffer (50mM glycylglycin, 5mM $MgCl_2$, 5mM ATP) and the "Relative Light Units" (RLU) were counted in a Lumat LB 9507 luminometer after addition of 100 ml of a 1 mM luciferin (Roche Diagnostics) solution. The mean value and standard deviation of the samples was calculated from the β -galactosidase and luciferase RLU values obtained from the four transfected wells of each sample.

C. Results: To set up an assay system for the measurement of C31-Int and Cre recombinase efficiency in mammalian cells the recombination substrate vector pRK64 shown in Figure 1E was first constructed. pRK64 contains a SV40 promoter region for expression in mammalian cells followed by a 1.1 kb cassette which consists of the coding region of the puromycin resistance gene and a polyadenylation signal sequence. This cassette is flanked at the 5'-end by the 84 bp attB and at the 3'-end by the 84 bp attP recognition site of C31-Int (Fig. 1 and 6). These attB and attP sites are located on the same DNA molecule and oriented in a way to each other which allows the deletion of the flanked DNA segment. The same orientation of attB and attP sites is used naturally by the Φ C31 phage and the bacterial genome, leading to the integration of the phage genome when both sites are located on different DNA molecules (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)). To measure C31-Int and Cre recombinase activities with the same substrate vector, pRK64 contains in addition two Cre recognition (loxP) sites in direct orientation next to the att sites. Since the att/lox-flanked cassette in plasmid pRK64 is inserted between the SV40 promoter and the coding region of the β -galactosidase gene, its presence inhibits

β -galactosidase expression as the SV40 promoter derived transcripts are terminated at the polyadenylation signal of the puromycin gene. Plasmid pRK64 is turned into a β -galactosidase expression vector upon C31-Int or Cre mediated deletion of the att/lox-flanked puromycin cassette since the remaining single att and loxP site do not substantially interfere with gene expression.

For the expression of recombinases a mammalian expression vector was designed which contains the CMV immediate early promoter followed by a hybrid intron, the coding region of the recombinase to be tested, and an artificial polyadenylation signal sequence. The backbone sequence of the four recombinase expression vectors shown in Figure 1A-D is identical to each other except for the recombinase coding region. Plasmid pCMV-C31Int(wt) (Fig. 1A) contains the nonmodified (wildtype) 1.85 kb coding region of C31-Int as found in the genome of phage Φ C31 (Kuhstoss, et al., J. Mol. Biol. 222, 897-908 (1991)). Plasmid pCMV-C31Int(NNLS) (Fig. 1B) contains a modified C31-Int gene coding for the full length C31-Int protein with a N-terminal extension of 7 amino acids derived from the SV40 virus large T antigen which serves as a nuclear localisation signal (NLS). Plasmid pCMV-C31Int(CNLS) (Fig. 1C) contains a C-terminal extension of 7 amino acids derived from the SV40 virus large T antigen which serves as a nuclear localisation signal (NLS). Plasmid pCMV-Cre (Fig. 1D) contains the 1.1 kb Cre coding region with an N-terminal fusion of the 7 amino acid NLS of the SV40 T-antigen. For Cre recombinase it has been shown that the N-terminal addition of the SV40 T-antigen NLS does not increase its recombination efficiency in mammalian cells (Le et al., Nucleic Acids Res., 27, 4703 - 4709 (1999)).

As a test system to compare the efficiency of the 4 recombinases the same amount of plasmid DNA of each of the recombinase expression vectors together with a fixed amount of the reporter plasmid pRK64 was transiently introduced into Chinese Hamster Ovary (CHO) cells. Thus, in this assay design the efficiency of the various recombinases on an extrachromosomal substrate introduced into the CHO cells was compared as a circular plasmid. Two days after transfection the cells from the various samples were lysed and the activity of β -galactosidase in the lysates was determined by a specific chemiluminescence assay and expressed in "Relative Light Units" (RLU (β -Gal)) (Fig. 2). In addition all samples

contained a fixed amount of a luciferase expression vector to control for the experimental variation of cell transfection and lysis. For this purpose the lysates of each sample were also tested for luciferase activity with a specific chemiluminescence assay and the values expressed as "Relative Light Units" (RLU (Luciferase)) (Fig. 2). All transfection samples contained in addition varying amounts of the unrelated cloning plasmid pUC19 so that all samples were equalised to the same amount of plasmid DNA. As a positive control for β -galactosidase a derivative of the recombination reporter pRK64 was used in which the loxP flanked 1.1 kb cassette has been removed through Cre mediated recombination in *E. coli* giving rise to plasmid pRK64(Δ Cre). As negative controls served samples which received the unrecombined reporter plasmid pRK64 but no recombinase expression vector as well as samples set up with the pUC19 plasmid alone.

To determine the relative efficiency of the tested recombinases the RLU values of β -galactosidase were divided individually for each sample by the RLU values obtained for luciferase and multiplied with 10^5 . From the values of the four data points of each sample the mean value and standard deviation was calculated as an indicator of recombinase activity (Gal/Luc) (Fig. 2). The relative activity of the tested recombinases was then compared to the positive control defined as an activity of 1.

As shown in Fig. 2, the expression of Cre recombinase (samples 10 and 11) resulted in a 150 to 170-fold increase of β -galactosidase activity as compared to the negative control (sample 3), demonstrating the wide dynamic range of our test system. Each recombinase vector was tested using two different amounts of DNA for transfection (0.5 and 1ng/sample), which in the case of Cre resulted in 63% and 72% recombinase activity (samples 10 and 11 as compared to the positive control). These two values establish that the DNA amounts used are close to the test systems saturation for recombinase expression as the doubling of DNA amounts resulted only in a minor increase of recombinase activity.

In comparison to Cre, the expression of wildtype C31-Int resulted in a considerably lower recombinase activity of 23% and 30% (Fig. 2, samples 4 and 5) as compared to the positive control. These values represent 37% and 42%

recombinase activity for wildtype C31-Int as compared to Cre recombinase (compare samples 4 and 5 with 10 and 11). Upon the expression of C31-Int fused with the N-terminal NLS (C31-Int(NNLS)) values of 32% and 36% recombinase activity (samples 6 and 7) were obtained as compared to the positive control. The C31-Int(NNLS) values represent 51% and 50% recombinase activity as compared to Cre (compare samples 6 and 7 to 10 and 11). Thus, the activity of C31-Int in mammalian cells is just moderately enhanced by the addition of a NLS signal.

Surprisingly, upon the expression of C31-Int fused with the C-terminal NLS (C31-Int(CNLS)) values of 50% and 65% recombinase activity (samples 8 and 9) were obtained as compared to the positive control. The C31-Int(CNLS) values represent 79% and 90% recombinase activity as compared to Cre recombinase (compare samples 8 and 9 to 10 and 11). Unexpectedly, C31-Int(CNLS) exhibits a dramatic, more than twofold increase of recombinase activity in comparison to C31-Int(wt) (compare samples 8 and 9 to 4 and 5).

In order to test whether the addition of a NLS sequence may be a general, simple method to enhance recombinase activity in mammalian cells we extended our studies by two additional recombinases: XisA recombinase (XisA) derived from the cyanobacterium *Anabaena*, and SSV-Integrase (SSV-Int) derived from the SSV1 virus of the thermophilic bacterium *Sulfolobus shibatae*. To this end we constructed mammalian expression vectors for the wildtype forms of XisA and SSV recombinases and compared their activity to versions which were modified by the N-terminal addition of the 7 amino acid NLS of the SV40 T-antigen. These recombinases were compared by the use of the reporter vector shown in Fig.1E, except that the att elements of C31-Int were replaced by the nif recognition sequences for XisA or the att sequences for SSV-Int. As described above for C31-Int, recombinase activities were tested by transient transfection into CHO cells using the reporter vector derived β -galactosidase activity as readout and cotransfected luciferase as internal control.

As shown in Fig.3 for both, XisA and SSV recombinases the addition of a NLS sequence did not improve their activity in a mammalian cell line as compared to the wildtype forms. At both DNA concentrations tested wildtype XisA exhibits a significant recombination activity as compared to the reporter vector alone (compare samples 2 and 3 to sample 1), but this activity is not altered by the

addition of an NLS (compare samples 2 and 3 to samples 4 and 5). SSV-Int exhibits only a low recombination activity (compare samples 7 and 8 with sample 6) which is also not enhanced by the addition of a NLS (compare samples 9 and 10 with samples 7 and 8). From these results we conclude that the addition of a
5 NLS to an inefficient recombinase is not a general, simple method to improve its performance in mammalian cells.

Taken together, in the transient transfection test system shown in Figure 2 a more than twofold activity increase of the Φ C31 Integrase could be achieved by
10 the C-terminal, but not the N-terminal addition of the SV40 T antigen NLS signal. As this signal sequence has been characterised to act as a nuclear localisation signal (Kalderon et. al, Cell, 39, 499 - 509 (1984)) we conclude that the efficiency increase of C31-Int(CNLS) is the result of the improved nuclear accumulation of this recombinase. The relative inefficiency of C31-Int (NNLS)
15 may be explained by the inaccessibility of the NLS peptide to the nuclear import machinery at the N-terminal position of the C31-Int protein.

In particular, it could be shown that C31-Int(CNLS) recombines extrachromosomal DNA in mammalian cells almost as efficient as the widely used Cre recombinase and thus provides an additional or alternative recombination
20 system of highest activity. The efficiency increase of C31-Int(CNLS) as compared to its wildtype form is regarded as an invention of substantial use for biotechnology.

Example 2

25 As demonstrated in example 1 C31-Int recombinase with the C-terminal fusion of the SV40 T-antigen NLS (C31-Int(CNLS)) shows in mammalian cells a recombination activity comparable to Cre recombinase on an extrachromosomal plasmid vector. It was further tried to test whether C31-Int(CNLS) exhibits a similar activity on a recombination substrate which is chromosomally integrated
30 into the genome of mammalian cells. This question is critical for the use of a recombination system for genome engineering as it is possible that a recombinase may act efficiently on extrachromosomal substrates but is impaired if the recognition sites are part of the mammalian chromatin. To characterise the recombination activity of C31-Int(CNLS) and C31-Int(NNLS) on a chromosomal
35 substrate the pRK64 reporter plasmid (Fig. 1E) was stably integrated, containing

a pair of loxP and att sites, into the genome of a mammalian cell line. One of the stable transfected clones was chosen for further analysis and was transiently transfected with recombinase expression vectors coding for C31-Int(CNLS), C31-Int(NNLS), C31-Int(wt) or Cre recombinase. The activity of β -galactosidase derived from the Cre expression vector recombined in these cells was taken as a measure of recombination efficiency.

A. Plasmid constructions: all plasmids used and their purification are described in example 1.

B. Cell culture and transfections: To generate a stably transfected C31-Int reporter cell line 2.5×10^6 NIH-3T3 cells (Andersson et al., Cell, 16, 63-75 (1979); DSMZ#ACC59; DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) were electroporated with 5 μ g pRK64 plasmid DNA linearised with ScaI and plated into 10cm petri dishes. The cells were grown in DMEM/Glutamax medium (Life Technologies) supplemented with 10% fetal calf serum at 37°C, 10% CO₂ in humid atmosphere, and passaged upon trypsinisation. Two days after tranfection the medium was supplemented with 1mg/ml of puromycin (Calbiochem) for the selection of stable integrants. Upon the growth of resistant colonies these were isolated under a stereomicroscope and individually expanded in the absence of puromycin. To demonstrate stable integration of the transfected vector, genomic DNA of puromycin resistant clones was prepared according to standard methods and 5-10 μ g were digested with EcoRV. Digested DNA was separated in a 0.8% agarose gel and transferred to nylon membranes (GeneScreen Plus, NEN DuPont) under alkaline conditions for 16 hours. The filter was dried and hybridised for 16 hours at 65°C with a probe representing the 5' part of the E. coli β -galactosidase gene (1.25 kb NotI - EcoRV fragment of plasmid CMV- β -pA (R. Kühn, unpublished). The probe was radiolabelled with P32-marked α -dCTP (Amersham) using the Megaprime Kit (Amersham). Hybridisation was performed in a buffer consisting of 10% dextranesulfate, 1% SDS, 50 mM Tris and 100 mM NaCl, pH7.5). After hybridisation the filter was washed with 2x SSC/1%SDS and exposed to BioMax MS1 X-ray films (Kodak) at - 80°C.

Transfections of the selected clone 3T3(pRK64)-3 with plasmid DNAs and the measurement of β -galactosidase activities in lysates were essentially performed

as described in example 1 for CHO cells, except that 32ng or 64ng of the recombinase expression plasmids and 218 or 186 ng of pUC19 plasmid were used and the pRK64 plasmid was omitted from all samples.

5 C. Histochemical detection of β -galactosidase activity in transfected 3T3(pRK64)-3 cells

To directly demonstrate β -galactosidase expression in recombinase transfected cells, 10^4 3T3(pRK64)-3 cells were plated one day before transfection into each well of a 48-well tissue culture plate (Falcon). For the transient transfection of
10 cells with plasmids each well received into 250 μ l of medium a total amount of 150 ng supercoiled plasmid DNA complexed before with the FuGene6 transfection reagent (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturers protocol. Each 150 ng DNA preparation contained 50 ng of the recombinase expression vector pCMV-Cre or pCMV-C31Int(CNLS) and 100ng of
15 the pUC19 plasmid. After 2 days the culture medium was removed from the wells, the wells were washed once with phosphate buffered saline (PBS), and the cells were fixed for 5 minutes at room temperature in a solution of 2% formaldehyde and 1% glutaraldehyde in PBS. Next, the cells were washed twice with PBS and finally incubated in X-Gal staining solution for 24 hours at 37°C
20 (staining solution: 5 mM $K_3(Fe(CN)_6)$, 5 mM $K_4(Fe(CN)_6)$, 2 mM $MgCl_2$, 1mg/ml X-Gal (BioMol) in PBS) until photographs were taken.

D. Results

To generate a mammalian cell clone with a stable genomic integration of the
25 C31-Int and Cre recombinase reporter plasmid pRK64, the murine fibroblast cell line NIH-3T3 was electroporated with linearised pRK64 DNA (Fig.1D; see also example 1) and subjected to selection in puromycin containing growth medium. Plasmid pRK64 contains in between the pair of loxP and att sites the coding region of the puromycin resistance gene expressed from the SV40-IE promoter.
30 Thirty-six puromycin resistant clones were isolated and the genomic DNA of 19 clones was analysed for the presence and copy number of the pRK64 DNA. Three clones, which apparently contain 2 - 4 copies of pRK64, were further characterised on the single cell level for the expression of β -galactosidase upon transient transfection with the Cre expression vector pCMV-Cre (Fig. 1C). The cell
35 clone with the largest proportion of β -galactosidase positive cells, 3T3(pRK64)-3,

was selected as most useful for the planned studies on C31-Int and Cre recombinase efficiency.

To compare the efficiency of wildtype C31-Int (C31-Int(wt)), C31-Int(CNLS),
5 C31-Int(NNLS), and Cre recombinases 32ng or 64 ng of the recombinase
expression vectors pCMV-C31Int(wt), pCMV-C31Int(CNLS), pCMV-C31Int(NNLS),
or pCMV-Cre (Fig. 1 A-D) together with the unrelated cloning plasmid pUC19
were transiently introduced into 3T3(pRK64)-3 cells, such that all samples
10 contained the same amount of plasmid DNA. As a negative control a sample
prepared with the pUC19 plasmid alone was used. Two days after transfection
the cells from the various samples were lysed and the activity of β -galactosidase
in the lysates was determined by a specific chemiluminescence assay and
expressed in "Relative Light Units" (RLU)(β -Gal) (Fig. 4). From the values of the
four data points of each sample the mean value and standard deviation was
15 calculated as an indicator of recombinase activity (Fig.4). The relative activity of
the tested recombinases was then compared to the highest value obtained with
the Cre expression vector, defined as an activity of 1.

As shown in Figure 4 the expression of Cre recombinase (samples 8 and 9)
20 resulted in a 36 to 49-fold increase of β -galactosidase activity as compared to the
negative control (sample 1), demonstrating the dynamic range of the test system
used. Each recombinase vector was tested using two different amounts of DNA
for transfection (32 ng and 64 ng/sample), which in the case of Cre resulted in
73% and 100% recombinase activity (samples 8 and 9). These two values
25 establish that the DNA amounts used are not far from the linear scale of the test
systems ability for recombinase expression as the twofold increase of the amount
of DNA also resulted in a significant increase of recombinase activity.

The expression of wildtype C31-Int (Fig. 4, samples 2 and 3) resulted in a low
30 recombinase activity of 4% and 10% as compared to the values obtained by Cre
transfection (compare samples 2 and 3 with 8 and 9). This activity was only
moderately enhanced by the expression of C31-Int(NNLS) to values of 19% and
22% of Cre activity (compare samples 4 and 5 with samples 8 and 9). Upon the
expression of C31-Int(CNLS) values of 48% and 78% recombinase activity were
35 obtained as compared to Cre recombinase (compare samples 6 and 7 to 8 and

9). Hence, C31-Int(CNLS) exhibits an 12-fold higher activity than C31-Int(wt) at 32 ng plasmid DNA (Fig.4, compare samples 6 and 2) and an 8-fold higher activity than C31-Int(wt) at 64 ng plasmid DNA (Fig.4, compare samples 7 and 3).

5 In addition, it was aimed to directly demonstrate in situ the expression of β -galactosidase in 3T3(pRK64)-3 cells after transfection with Cre or C31-Int(CNLS) recombinase plasmid. Two days after transfection the cells were fixed in situ and incubated with the histochemical X-Gal assay which detects β -galactosidase
10 positive cells by a blue precipitate. As shown in Figure 5 stained cells were found at a comparable frequency in the samples transfected with the Cre or C31-Int(CNLS) expression vectors but not in the nontransfected control. This result confirms that the β -galactosidase activities measured by chemiluminescence upon recombinase transfection (Fig. 4) results from a population of individual,
15 recombined reporter cells.

In conclusion, upon the transient transfection of recombinase expression vectors into a cell line with a genomic integration of the recombination substrate vector, a 8 – 12-fold activity increase of the Φ C31 Integrase by the C-terminal fusion
20 with the SV40 T-antigen NLS signal was found. As this signal sequence has been characterised to act as a nuclear localisation signal (Kalderon et. al, Cell, 39, 499 – 509 (1984)), it was concluded that the dramatic efficiency increase of C31-Int(CNLS) is the result of the improved nuclear accumulation of this recombinase. The approximately tenfold activity increase of C31-Int(CNLS) upon
25 expression in a cell line with a genomic integration of the substrate vector is considerably higher than the activity increase found upon the transient expression of both vectors (see example 1). Thus, a substrate vector integrated into the chromatin of a mammalian cell may pose more stringent requirements on recombinase activity to be recombined as compared to an extrachromosomal
30 substrate.

The dramatic activity increase of C31-Int(CNLS), as compared to its wildtype form, on a stable integrated substrate in mammalian cells is an invention of significant practical use as this recombinase is as efficient as the widely used

Cre/loxP system; thus, C31-Int(CNLS) provides an additional or alternative recombination system of highest activity.

Example 3

5 To demonstrate that the increase in β -galactosidase activity obtained by the cotransfection of a C31-Int expression vector and the reporter vector pRK64 into mammalian cells is in fact the result of recombinase mediated deletion, one of the recombination products was detected by a specific polymerase chain reaction (PCR). The amplified PCR product was cloned and its sequence determined. The
10 obtained sequence confirms that C31-Int mediated deletion of the test vector occurs in a mammalian cell line and that the recombination occurs at the known breakpoint within the attB and attP sites.

A. Plasmid constructions: The construction of plasmids pRK64, pCMV-Cre and
15 pCMV-C31-Int(wt) is described in Example 1. To simulate the recombination of pRK64 by C31-Int, the sequence between the CAA motives of the att sites (boxed in Fig.5) was deleted from the computerfile of pRK64, giving rise to the sequence of pRK64(Δ Int) (SEQ ID NO:16).

20 B. Transfection of Cells and PCR amplification: MEF5-5 mouse fibroblasts (Schwenk et al., 1998) (20000 cells per well of a 12 well plate (Falcon)) were transfected with 0.5 μ g pRK64 alone or together with 250 ng pCMV-Int(wt) or pCMV-Cre using the FuGene6 transfection reagent following the manufacturers protocol (Roche Diagnostics). After 2 days DNA was extracted from these cells
25 according to standard methods and used for PCR amplification with Primers P64-1 (SEQ ID NO:17; complementary to position 111-135 of pRK64(Δ Int)) and P64-4 (SEQ ID NO:18; complementary to position 740-714 of pRK64(Δ Int)) using the Expand High Fidelity PCR kit (Roche Diagnostics). PCR products were separated on a 0.8% agarose gel, extracted with the QuiaEx kit (Quiagen) and cloned into
30 the pCR2.1 vector using the TA cloning kit (Invitrogen) resulting in plasmid pRK80d. The sequence of its insert, seq80d (SEQ ID NO:19), was determined using the reverse sequencing primer and standard sequencing methods (MWG Biotech).

35 For the measurement of β -galactosidase activity the cells were lysed 2 days after

transfection and the β -galactosidase activities were determined with the β -galactosidase reporter gene assay (Roche Diagnostics) as described in example 1.

5 C. Results: As a test vector for C31-Int mediated DNA recombination plasmid pRK64 was used, which contains the 1.1 kb coding region of the puromycin resistance gene flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site of C31-Int (Fig. 5). These attB and attP sites are located on the same DNA molecule and oriented in a way to each other which allows the
10 deletion of the att-flanked DNA segment. The same orientation of attB and attP sites is used naturally by the Φ C31 phage and the bacterial genome for the integration of the phage genome when both sites are located on different DNA molecules (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)). As a positive control, vector pRK64 contains in addition two Cre recombinase
15 recognition (loxP) sites in direct orientation next to the att sites. Since the att-flanked DNA segment in plasmid pRK64 is inserted between a promoter active in mammalian cells and the β -galactosidase gene, its deletion can be measured by the increase of β -galactosidase activity. The expected product of C31-Int mediated deletion of plasmid pRK64 is shown in Fig. 6, designated as
20 pRK64(Δ Int). If the recombination between attB and attP occurs as described in bacteria (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)), a single attR site is generated and left on the parental plasmid (Fig. 6) while the flanked DNA is excised and contains an attL site. Beside the measurement of β -galactosidase activity, C31-Int mediated recombination of pRK64 can be directly
25 detected on the DNA level by a specific polymerase chain reaction (PCR) using the primers P64-1 and P64-4 (Fig. 6). These primers, located 5' of the attB site (P64-1) and 3' of the attP site, are designed to amplify a PCR product of 630 bp length upon the C31-Int mediated recombination of pRK64. For the expression of C31-Int in mammalian cells plasmid pCMV-C31(wt) was used, which contains the
30 CMV-IE-Promoter upstream of the C31-Int coding region followed by a synthetic polyadenylation signal (see Example 1 and Fig.1).

The recombination substrate vector pRK64 was transiently transfected into the murine fibroblast cell line MEF5-5 either alone, or together with the C31-Int expression vector pCMV-C31(wt), or together with an expression vector for Cre
35 recombinase, pCMV-Cre. Two days after transfection half the cells of each sample

was lysed and used to measure β -galactosidase activity by chemiluminescence, and the other half was used for the preparation of DNA from the transfected cells for PCR analysis. The β -galactosidase levels of the 3 samples were found as following (expressed as "Relative Light Units" (RLU) with standard deviation (SD) of the β -galactosidase assay):

Sample	RLU (SD)
1) pRK64	692 \pm 5
2) pRK64 + pCMV-Cre	8527 \pm 269
3) pRK64 + pCMV-C31(wt)	1288 \pm 93

As the coexpression of the test vector pRK64 together with the C31-Int expression vector in sample 3 leads to a significant increase of β -galactosidase activity as compared to pRK64 alone, this result suggests that pRK64 is recombined by C31-Int as anticipated in Fig. 6.

Next, cellular DNA was prepared from the three samples and tested for the occurrence of the expected Cre or C31-Int generated deletion product by PCR using primers P64-1 and P64-4 for amplification. As shown in Fig. 7 an amplification product of the expected size was found only in the samples cotransfected with the Cre or C31-Int recombinase expression vectors (Fig. 7A, lane3 and lane 4). The PCR products amplified from pRK64 recombined by C31-Int or Cre are of the same size but should be recombined via the attB/P or loxP sites, respectively.

To prove that the PCR product found after cotransfection of plasmids pRK64 and pCMV-C31(wt) represents in fact the deletion product of C31-Int mediated recombination, this DNA fragment was cloned into a plasmid vector and its DNA sequence determined. One clone, pRK80d, was analysed, and its sequence showed exactly the sequence of an attR site as expected from C31-Int mediated deletion of pRK64 (Fig. 7B, compare to Fig. 6).

In conclusion, this experiment demonstrates that C31-Int mediated deletion of a vector containing a pair of attB/attP sites occurs in a mammalian cell line, and that the recombination occurs within the same 3 bp breakpoint region of attB and attP as found in bacteria (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 -

5510 (1998)). Thus, it was concluded that an increase of β -galactosidase activity observed by cotransfection of the pRK64 reporter vector and a C31-Int expression vector in mammalian cells truly reflects C31-Int recombinase activity.

5

Example 4

As has been demonstrated in examples 1-3, the C31-Int recombinase with the C-terminal fusion of the SV40 T-antigen NLS (C31-Int(CNLS)) shows a recombination activity comparable to Cre recombinase on an extrachromosomal as well as a chromosomally integrated target in mammalian cells in vitro. To test whether C31-Int(CNLS) exhibits activity in mice, transgenic mice carrying a C31-Int(CNLS) expression vector were generated. These transgenic mice were crossed with reporter mice carrying the recombinase substrate. Recombination-mediated expression of β -galactosidase, which can be measured by staining with the substrate X-Gal, was analyzed in testes of double transgenic progeny carrying both the recombinase and the reporter.

A. Plasmid constructions: For the construction of the C31-Int(CNLS) transgene expression vector, the C31Int gene with C-terminal NLS was isolated as a 2 kb-fragment generated by restriction of pCMV-C31Int(CNLS) (SEQ ID NO: 12) with BglII. The fragment was ligated into the BglII restriction site of the vector pCAGGS-Cre-pA (SEQ ID NO:104) giving rise to the plasmid pCAGGS-C31CNLS-pA (SEQ ID NO:105). In pCAGGS-C31CNLS-pA the C31-Int(CNLS) (position 1891-3753) is transcribed from the CAGGS promoter (position 1-1616) and followed by the SV40 late region polyadenylation sequence (position 3763-3941).

B. Production of transgenic mice: For the embryo injections a 3.95 kb-fragment was generated by restriction of the plasmid pCAGGS-C31CNLS-pA with PstI and AscI. This fragment was purified as follows: DNA bands were separated on an agarose-gel without ethidiumbromide. One part of the gel was stained with ethidiumbromide to locate the band to excise. The DNA was electroeluted from the excised band with S&S Biotrap Elution Chamber in 1x TAE (40 mM Tris-acetate, 1 mM EDTA) overnight. The DNA was precipitated from the eluate with 1/10 volume 3M sodium acetate and 2.5 volumes ethanol at -20 °C for several

hours. The DNA was pelleted by centrifugation at 13000 rpm for 30 min and washed twice with 70 % ethanol. The dried DNA pellet was resuspended in TE (10 mM Tris, 1 mM EDTA, pH 8). Subsequently the precipitation procedure was repeated once and the DNA resuspended in injection buffer (10 mM Tris pH 7.2, 0.1 mM EDTA). The sample was dialysed with Slide-A-Lyse Mini Dialysis Unit (Pierce) in injection buffer with several changes of buffer at 4°C overnight. Different amounts of the sample were checked on a gel to determine concentration. To generate transgenic mice, 5-10 fg of the purified fragment were injected into one pronucleus of (B6CBA)F2 mouse one-cell embryos. The injected embryos were subsequently transferred into the oviduct of 0.5 day pseudopregnant NMRI females.

C. Analysis of transgenic mice: Mice were analyzed for the presence of the pCAGGS-C31CNLS-pA transgene by PCR using tail DNA and the primers C31-screen 1 (SEQ ID NO:100) and C31-screen 2 (SEQ ID NO:101) amplifying a fragment of 500 bp. The PCR reaction contained 5 µl PCR buffer (Invitrogen), 2 µl 50 mM MgCl₂, 1.5 µl 10 mM dNTP-mix, 2 µl (10 pmol) of each primer, 0.5 µl Taq-polymerase (5 U/ µl) and water to a volume of 50 µl. The program used for the PCR reactions was: 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min in 30 cycles.

D. Analysis of C31-Int(CNLS) activity: Founder mice transgenic for the pCAGGS-C31CNLS-pA transgene were crossed to heterozygous C31 reporter mice carrying the C31 reporter construct in the ROSA26 locus (SEQ ID NO:106) (Fig. 8). Offspring of the crosses were genotyped for the presence of the pCAGGS-C31CNLS-pA transgene by the PCR assay described in section C as well as for the ROSA26-C31 reporter allele by a LacZ-specific PCR assay. The PCR was performed using tail DNA and the primers β-Gal 3 (SEQ ID NO:102) and β-Gal 4 (SEQ ID NO:103) amplifying a fragment of 315 bp. The PCR reaction contained 5 µl PCR buffer (Invitrogen), 2.5 µl 50 mM MgCl₂, 2 µl 10 mM dNTP-mix, 1 µl (10 pmol) of each primer, 0.4 µl Taq-polymerase (5 U/ µl) and water to a volume of 50 µl. The program used for the PCR reactions was: 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1 min in 30 cycles.

Testes from mice carrying the pCAGGS-C31CNLS-pA transgene as well as the reporter locus and from a control mouse carrying the reporter allele only were

dissected. The tissues were imbedded in OCT Tissue freezing medium (Leica/Jung) and frozen in liquid nitrogen. Cryosections were generated from the embedded tissues using a Leica CM3050 cryomicrotome, dried on polylysine-coated slides for 1-4 hours and then stained as follows: Sections were fixed in 0.2 % glutaraldehyde, 5 mM EGTA, 2 mM MgCl₂ in 0.1 M PB (K₂HPO₄/ KH₂PO₄, pH 7.3) for 5 min at room temperature and washed in wash buffer (2 mM MgCl₂, 0.02 % Nonidet-40 in PB in 0.1 M PB) 3 times for 15 min. Then sections were stained in X-Gal-solution (0.6 mg/ ml X-Gal in DMSO, 5 mM potassium hexacyanoferrat III, 5 mM potassium hexacyanoferrat II in LacZ wash buffer) overnight at 37 °C. After staining sections were washed in 1x PBS twice for 5 min. Dehydration was performed by washing the sections first with 70 %, 96 % and 100 % ethanol for 2 min each, then with a 1:1 mix of ethanol and xylol for 5 min and in the end only with xylol for 5 min. Before taking pictures sections were mounted in Entellan.

E. Results: To identify transgenic founder mice carrying the pCAGGS-C31CNLS-pA transgene, 29 mice born from the injection experiment were analyzed for the presence of the transgene. 5 founder mice (3 females and 2 males) were identified. To analyze the activity of the C31-Int(CNLS) recombinase in transgenic mice, 2 of the female founder mice were crossed to heterozygous C31 reporter mice carrying a C31 reporter construct in the ROSA26 locus (Fig. 8). From each of these crosses, one offspring carrying the pCAGGS-C31CNLS-pA transgene as well as the C31 reporter allele was sacrificed. In order to determine whether pCAGGS-C31CNLS-pA transgenic mice are able to delete an attB/P flanked DNA sequence in the mouse germline, tissue sections from the testes of the sacrificed animals were prepared and stained for β -galactosidase activity with X-Gal. Fig. 9 shows the result of the staining experiment for one of these mice (A) as well as a control mouse carrying only the reporter allele, but lacking the pCAGGS-C31CNLS-pA transgene (B). Clear staining can be detected in the maturing sperm cells in about 50% of the tubules with the proportion of β -galactosidase expressing cells ranging between 10 and 100. No staining could be detected for the control mouse. This clearly demonstrates that C31-int-mediated recombination has taken place during spermatogenesis in the pCAGGS-C31CNLS-pA transgenic mice. These results show that the C31-int is functional *in vivo*, in a

transgenic mouse system and therefore provides a new tool to introduce specific deletions, inversions or integrations into the mouse germline.

Claims

1. A fusion protein comprising
 - (a) a recombinase domain comprising a recombinase protein or or a mutant thereof having a recombinase activity similar to that of the corresponding wild-type recombinase and
 - (b) a signal peptide domain linked to said recombinase domain which directs nuclear import of said fusion protein in eucaryotic cells.
2. The fusion protein of claim 1, wherein the activity of the fusion protein in eucaryotic cells is significantly higher as compared to that of the wild-type recombinase corresponding to the recombinase of the recombinase domain.
3. The fusion protein of claim 1 or 2, wherein the recombinase domain comprises a recombinase protein belonging to the family of large serine recombinases or a mutant thereof, preferably the recombinase domain comprises a recombinase protein selected from the group consisting of bacteriophage Φ C31 integrase (C31-Int), coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase, CisA recombinase, XisF recombinase, transposon Tn4451 TnpX recombinase and lactococcal bacteriophage TP901-1 recombinase, or a mutant thereof; most preferably the recombinase protein is a C31-Int protein or a mutant thereof.
4. The fusion protein of claim 3, wherein the recombinase protein comprises a C31-Int having the amino acid sequence shown in SEQ ID NO:21 or a C-terminal truncated form thereof, said truncated form of the C31-Int preferably comprising amino acid residues of 306 to 613 of SEQ ID NO:21.
5. The fusion protein according to any one of claims 1 to 4, wherein the signal peptide domain is derived from yeast GAL4, SKI3, L29 or histone H2B proteins, polyoma virus large T protein, VP1 or VP2 capsid protein, SV40 VP1 or VP2 capsid protein, adenovirus E1a or DBP protein, influenza virus NS1 protein, hepatitis virus core antigen or the mammalian lamin, c-myc, max, c-myb, p53, c-erbA, jun, Tax, steroid receptor or Mx proteins, SV40 T-antigen or other proteins with

known nuclear localisation, preferably the signal peptide domain comprises a peptide which is derived from the SV40 T-antigen.

6. The fusion protein according to any one of claims 1 to 5, wherein the signal peptide domain

(i) has a length of 5 to 74, preferably 7 to 15 amino acid residues, and/or
(ii) comprises a segment of 6 amino acid residues having at least 2 positively charged basic amino acid residues, said basic amino acid residues being preferably selected from lysine, arginine and histidine.

7. The fusion protein of claim 5 or 6, wherein the signal peptide domain comprises a peptide selected from a sequence shown in SEQ ID NOs:24 to 53, preferably the signal peptide comprises the amino acid sequence Pro-Lys-Lys-Lys-Arg-Lys-Val (SEQ ID NO:53).

8. The fusion protein according to any one of claims 1 to 6, wherein

(i) the signal peptide domain is linked to the N-terminal or C-terminal of the recombinase domain or is integrated into the recombinase domain, preferably the signal peptide domain is linked to the C-terminal of the recombinase domain; and/or

(ii) the signal peptide domain is linked to the recombinase domain directly or through a linker peptide, said linker preferably having 1 to 30 essentially neutral amino acid residues.

9. The fusion protein of claim 1 comprising the amino acid sequence shown in SEQ ID NO:23.

10. A DNA coding for the fusion protein according to any one of claims 1 to 9.

11. A vector containing the DNA as defined in claim 10.

12. A microorganism containing the DNA of claim 10 and/or the vector of claim 11.

13. A process for preparing the fusion protein as defined in any one of claims 1 to 9 which comprises culturing a microorganism as defined in claim 11 under conditions suitable for expression of said fusion protein and recovering said fusion protein.

5

14. Use of the fusion protein as defined in any one of claims 1 to 9 to recombine DNA molecules, which contain recombinase recognition sequences for the recombinase protein of the recombinase domain, in eucaryotic cells.

10 15. A cell, preferably a mammalian cell containing the DNA sequence of claim 10 in its genome.

16. The cell of claim 15, also containing recognition sequences for the recombinase protein of the recombinase domain in its genome.

15

17. Use of the cell of claim 15 or 16 for studying the function of genes and for the creation of transgenic organisms.

18. A transgenic organism, preferably a transgenic non-human mammal
20 containing the DNA sequence of claim 10 in its genome.

19. The transgenic organism of claim 18 also containing recognition sequences for the recombinase protein of the recombinase domain in its genome.

25 20. Use of the transgenic organism of claim 18 or 19 for studying gene function at various developmental stages.

21. A method for recombining DNA molecules of cells or organisms containing recombinase recognition sequences for the recombinase protein of the
30 recombinase domain as defined in claims 1 to 9, which method comprises supplying the cells or organisms with a fusion protein as defined in claims 1 to 9 or with a DNA sequence of claim 10 and/or a vector of claim 11 which are capable of expressing said fusion protein in the cell or organism.

35 22. A method for recombining a DNA molecule containing recognition sequences

for a recombinase protein in a eucaryotic cell, said method comprising contacting the cell with a fusion protein according to claim 1 that recognizes said recognition sequences, wherein the fusion protein catalyzes recombination of the DNA molecule.

5

23. The fusion protein according to any one of claims 1 to 9 which catalyzes recombination at recognition sequences for the recombinase protein.

24 A transgenic organism, preferably a transgenic non-human mammal,

10

comprising a cell containing a DNA sequence coding for a recombinase fusion protein as defined in claims 1 to 9 and 23 in its genome.

SEQUENCE LISTING

5 <110> Artemis Pharmaceuticals GmbH
 <120> Modified Recombinase
 <130> 012787wo/JH/ml
 10 <140>
 <141>
 <160> 108
 15 <170> PatentIn Ver. 2.1
 <210> 1
 <211> 86
 <212> DNA
 20 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: primer C31-4
 25 <400> 1
 cgtgacggtc tcgaagccgc ggtgcggggtg ccagggcggtg cccttgggct ccccgggcgc 60
 gtactccacc tcacccatct ggtcca 86
 30 <210> 2
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 35 <220>
 <223> Description of Artificial Sequence: primer C31-5
 <400> 2
 40 cgtggaccag atgggtgagg tggagtacgc gcccggggag cccaagggca cgccctggcc 60
 cacgcaccgc ggcttcgaga ccgtca 86
 <210> 3
 <211> 90
 45 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: primer C31-6
 50 <400> 3
 gatccagaag cggttttcgg gagtagtgcc ccaactgggg taacctttga gttctctcag 60
 ttgggggcgt aggtcgcgc acatgacacg 90
 55 <210> 4
 <211> 90
 <212> DNA
 <213> Artificial Sequence
 60 <220>
 <223> Description of Artificial Sequence: primer C31-7-2
 <400> 4
 65 gatccgtgtc atgtcggcga cctacgccc ccaactgaga gaactcaaag gttaccccag 60
 ttggggcact actcccga aa accgttctg 90

<210> 5
 <211> 7438
 <212> DNA

5 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector pRK64

10 <400> 5

```

cgctcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagttaggg tgtggaaagt 60
ccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag tcagcaacca 120
ggctcccccag caggcagaag tgtgcaaaag atgcatctca attagtcagc aaccatagtc 180
ccgcccctaa ctccgcccct cccgccccta actccgcccga gttccgcccga ttctccgccc 240
catggctgac taatTTTTTT tatTTatgca gaggcgagg ccgcctcggc ctaggaaacag 300
tcgaagacac tgcagagacc tacttcacta acaaccggta cagttcgtgg accagatggg 360
tgaggtggag tacgcgcccg gggagcccga gggcagccc tggcaccgcg accgcggctt 420
cgagaccgtc acgaataact tcgtatagca tacattatac gaagttataa gcttgcctgc 480
ctgcaggctg ccgcccacga ccggccggcc ggtgcgccc caatcccctg acccagccc 540
ctgaccccctc acaaggagac gaccttccat gaccgagtac aagcccacgg tgcgcctcgc 600
caccgcgcgac gacgtccccc gggcgcgtac caccctcgcc gccgcgttcg ccgactaccc 660
cgccacgcgc caccacgtcg acccggaacc ccacatcgag cgggtcaccc agctgcaaga 720
actcttccctc acgcgcgtcg ggctcgacat cggcaagggt tgggtcgccg acgacggcgc 780
cgcggtggcg gtctggacca cgcggagag cgtcgaagcg ggggcgggtg tcgccgagat 840
cggcccgcg atggccgagt tgagcgggtc ccggtggcc gcgcagcaac agatggaagg 900
cctcctggcg ccgcaccggc ccaaggagcc cgcgtgggtc ctggccaccg tcgcgctctc 960
gcccgaccac cagggcaagg gtctgggcag cgcgctcgtg ctcccggag tggaggcggc 1020
cgagcgcgcg ggggtgcccg ccttcctgga gacctccgcg ccccgcaacc tcccctctca 1080
cgagcggctc ggcttcaccg tcaccgcccga cgtcagtgcc ccgaaggacc gcgcgacctg 1140
gtgcatgacc cgcaagcccg gtgcctgacg cccgcccac gaccgcagc gcccgaccga 1200
aaggagcgca cgaccccctg gctccgaccg aagccgaccc gggcgccccc gccgacccc 1260
caccgcgcgc cgaggcccac cgaactctaga ggatcataat cagccatacc acattttagt 1320
aggttttact tgcttttaaa aacctcccac acctcccct gaacctgaaa cataaatga 1380
atgcaattgt tgttgTTaac ttgtttattg cagcttataa tggttacaaa taaagcaata 1440
35 gcatcacaaa tttcacaaat aaagcatttt ttctactgca ttctagtTgt ggtttgtcca 1500
aactcctcaa tgtatttat catgtctgga tccgtgtcat gtcggcgacc ctacgcccc 1560
aactgagaga actcaaaagg taocccaggt gggcgactac tcccgaaaac cgcttctgga 1620
tccataactt cgtatagcat acattatacg aagttatacc gggccaccat ggtcgcgagt 1680
agcttggaac tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg cgttacccaa 1740
40 cttaatcgcc ttgaagcaca tcccctttc gccagctggc gtaatagcga agaggccgc 1800
accgatcgcc cttcccacaa gttgcgcagc ctgaatggcg aatggcgctt tgccctggtt 1860
ccggcaccgg aagcggcgcc ggaaagctgg ctggagtgcg atcttccctga ggcgatact 1920
gtcgtcgtcc cctcaaaact gcagatgcac ggttacgatg cgcctatcta caccacgta 1980
acctatccca ttacggtcaa tccgcgcttt gttcccacgg agaatccgac gggttgttac 2040
45 tcgctcacat ttaatgttga tgaaagctgg ctacaggaag gccagacggc aattattttt 2100
gatggcgTTa actcggcggt tcatctgtgg tgcaacgggc gctgggtcgg ttacggccag 2160
gacagtctgt tgcgctcgt atttgacct agcgcatttt tacgcgcgg agaaaaccgc 2220
ctcgcggtga ttggtcgtcg ttggagtgcac ggcagttatc tggaagatca gcatctgtg 2280
cggatgagcg gcattttccg tgacgtctcg ttgctgcata aaccgactac acaaatcagc 2340
50 gatttccatg ttgccactcg ctttaatgat gatttcagcc gcgctgtact ggaggctgaa 2400
gttcagatgt gcggcgagtt gcgtgactac ctacgggtaa cagtttcttt atggcagggt 2460
gaaacgcagg tcgccagcgg caccgcgcct ttccggcggt aaattatcga tgagcgtggt 2520
ggttatgcgc actcgcgtc acactcgtct aacgtcgaaa acccgaaact gtggagcgcc 2580
55 gaaatcccga atctctatcg tcgggtggtt gaactgcaca ccgcgcagcg cacgctgatt 2640
gaagcagaag cctgcgatgt cggtttccgc gaggtgcgga ttgaaaatgg tctgctgctg 2700
ctgaacggca agcggttgct gattcgaggg gttaacgcgt acgagcatca tctctgcat 2760
ggtcaggtca tggatgagca gacgatggtg caggatatcc tgctgatgaa gcagaacaac 2820
tttaacgcgc tgcgctgttc gcattatccg aaccatccgc tgtggtacac gctgtgcgac 2880
60 cgctacggcc tgtatgtgtt ggtatgaagg aatatgaaa cccacggcat ggtgccaatg 2940
aatcgtctga ccgatgatcc gcgctggcta ccggcgatga gcgaacgcgt aacgcgaatg 3000
gtgcagcgcg atcgtaatca cccgagtgtg atcatctggt cgtgggggaa tgaatcaggc 3060
cacggcgcta atcagcagc gctgtatcgc tggatcaaat ctgtcgatcc ttccgccc 3120
gtgcagtatg aaggcggcgg agccgacacc acggccaccg atattatttg cccgatgtac 3180
65 gcgcgcgtgg atgaagacca gcccttcccc gctgtgccga aatggtccat caaaaaatgg 3240
ctttcgctac cgtgagagac gcgcccgctg atcctttgcg aatacgcca cgcgatgggt 3300
aacagtcctg cgggtttcgc taaatactgg caggctttc gtcagtatcc ccgtttacag 3360
ggcggcttcg tctgggactg ggtggatcag tcgctgatta aatatgatga aaacggcaac 3420

```

	ccgtgggtcgg	cttacggcgg	tgatTTTTggc	gatagccgga	acgatcgcca	gttctgtatg	3480
	aacgggtctgg	totttggcga	cgcacgcgg	catccagcgc	tgacggaagc	aaaacaccag	3540
	cagcagttttt	tccagttccg	tttatccggg	caaaccatcg	aagtgaccag	cgaataacctg	3600
5	ttccgtcata	gcgataacga	gctcctgcac	tggtatgggtg	cgctggatgg	taagccgctg	3660
	gcaagcgggtg	aagtgcctct	ggatgtcgct	ccacaaggta	aacagttgat	tgaactgcct	3720
	gaactaccgc	agccggagag	cgccggggcaa	ctctggctca	cagtacgcgt	agtgaaccg	3780
	aacgcgacgg	catggtcaga	agccggggcac	atcagcgctt	ggcagcagtg	gcgtctggcg	3840
	gaaaacctca	gtgtgacgct	ccccggcgcg	tcccacgcca	tcccgcatct	gaocaccagc	3900
10	gaaatggatt	tttgcacoga	gctgggtaat	aagcgttggc	aatttaaccg	ccagtcaggc	3960
	tttctttcac	agatgtggat	tggcgataaa	aaacaactgc	tgacgcccgt	gcgcgatcag	4020
	ttcacccttg	caccgctgga	taacgacatt	ggcgtaagtg	aagcgacccg	cattgaccct	4080
	aacgcctggg	togaacgctg	gaaggcggcg	ggccattacc	aggccgaagc	agcgttgttg	4140
	cagtgcacgg	cagatacact	tgctgatgcg	gtgctgatta	cgaccgctca	cgogtggcag	4200
15	catcaggggga	aaaccttatt	tatcagccgg	aaaacctacc	ggattgatgg	tagtgggtcaa	4260
	atggcgatta	ccgttgatgt	tgaagtggcg	agcgatacac	cgcatccggc	gcggattggc	4320
	ctgaactgcc	agctggcgca	ggtagcagag	cgggtaaaact	ggctcggatt	agggccgcaa	4380
	gaaaactatc	ccgaccgcct	tactgcccgc	tgTTTTgacc	gctgggatct	gccattgtca	4440
	gacatgtata	ccccgtacgt	cttcccggagc	gaaaacggtc	tgcgctgcgg	gacgcgcgaa	4500
20	ttgaattatg	gtccacacca	gtggcgggcg	gacttccagt	tcaacatcag	ccgctacagt	4560
	caacagcaac	tgatggaaac	cagccatcgc	catctgctgc	acgcggaaga	aggcacatgg	4620
	ctgaatatcg	acggtttcca	tatggggatt	ggtggcgacg	actcctggag	cccgctcagta	4680
	tccggcggaat	tccagctgag	cgccgggtcgc	taccattacc	agttgggtctg	gtgtcaaaaa	4740
	taataataac	cgggcagggg	ggatctttgt	gaaggaaact	tacttctgtg	gtgtgacata	4800
25	attggacaaa	ctacctacag	agattttaaag	ctctaaggta	aataataaaat	ttttaagtgt	4860
	ataatgtgtt	aaactactga	ttctaattgt	tttgttattt	tagattccaa	cctatggaaac	4920
	tgatgaatgg	gagcagtggt	ggaatgccag	atccagacat	gataagatac	attgatgagt	4980
	ttggacaaac	cacaactaga	atgcagtga	aaaaatgctt	tatttgtgaa	atttgtgatg	5040
	ctattgcttt	atttgttaacc	attataagct	gcaataaaca	agttaacaac	aacaattgca	5100
30	ttcattttat	gtttcagggt	cagggggagg	tgtgggagg	tttttaaagc	aagtaaaacc	5160
	tctacaaatg	tggtatggct	gattatgatc	tgcggccgca	gggcctcgtg	atacgccatg	5220
	ttttataggt	taatgtcatg	ataataatgg	tttcttagac	gtcagggtggc	acttttcggg	5280
	gaaatgtgcg	cggaacccct	atttgtttat	ttttctaaat	acattcaaat	atgtatccgc	5340
	tcatgagaca	ataacccctga	taaatgcttc	aataatattg	aaaaaggaag	agtatgagta	5400
35	ttcaacattt	ccgtgtcgcc	cttattccct	tttttgccgc	attttgcctt	cctgtttttg	5460
	ctcacccgga	aacgctgggt	aaagtaaaag	atgctgaaga	tcagttgggt	gcacgagtg	5520
	gttacatcga	actggatctc	aacagcggta	agatccttga	gagttttcgc	cccgagaagc	5580
	gttttccaat	gatgagcaat	tttaaaagtt	tgctatgtgg	cgcggtatta	tcccttattg	5640
	acgcggggca	agagcaactc	ggtcgccgca	tacactattc	tcagaatgac	ttggttgagt	5700
40	actcaccagt	cacagaaaag	catcttacgg	atggcatgac	agtaagagaa	ttatgcagtg	5760
	ctgccataac	catgagtgat	aacactgcgg	ccaacttact	tctgacaacg	atcggaggac	5820
	cgaaggagct	aaccgctttt	ttgcacaaca	tgggggatca	tgtaactcgc	cttgatcggt	5880
	gggaaccggg	gctgaatgaa	gocataccaa	acagcagcgc	tgacaccacg	atgcctgtag	5940
	caatggcaac	aacgttgccg	aaactattaa	ctggcgaaact	acttactcta	gcttccggcg	6000
45	aacaattaat	agactggatg	gaggcggata	aagttgcagg	accacttctg	cgctcggccc	6060
	ttccggctgg	ctggtttatt	gotgataaat	ctggagccgg	tgagcgtggg	tctcgggta	6120
	tcattgcagc	actggggcca	gatggtaagc	cctcccgat	cgtagttatc	tacacgacgg	6180
	ggagtcaggc	aactatggat	gaacgaata	gacagatcgc	tgagataggt	gcctcactga	6240
	ttaacgattg	gtaactgtca	gaccaagttt	actcatatat	actttagatt	gatttgaatac	6300
50	ttcattttta	attttaaagg	atctaggtga	agatcccttt	tgataatctc	atgacaaaaa	6360
	tcccttaacg	tgagttttcg	ttccactgag	cgtcagaccc	cgtagaaaag	atcaaaggat	6420
	cttcttgaga	tccttttttt	ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccacccg	6480
	taccagcgg	ggtttgtttg	coggatcaag	agctaccaac	tctttttccg	aaggtaactg	6540
	gcttcagcag	agcgcagata	ccaaatactg	tccctctagt	gtagccgtag	ttaggccacc	6600
55	acttcaagaa	ctctgtagca	cgcctacat	acctcgctct	gctaactcctg	ttaccagtgg	6660
	ctgctgccag	tggcgataag	togtgtctta	ccgggttggg	ctcaagacga	tagttaccgg	6720
	ataaggcgca	gcggctgggc	tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	6780
	cgaactacac	cgaactgaga	taoctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	6840
	aaggggagaaa	ggcggacagg	tatccggtaa	gcggcagggt	cggaacagga	gagcgcacga	6900
60	gggagcttcc	agggggaaac	gcctgggtatc	tttatagtc	tgctggggtt	cgccacctct	6960
	gacttgagcg	ctgatttttg	tgatgctcgt	cagggggggc	gagcctatgg	aaaaacgcca	7020
	gcaacgcggc	ctttttacgg	ttcctggcct	tttgctggcc	ttttgctcac	atgttctttc	7080
	ctgcgttatc	ccctgattct	gtggataacc	gtattaccgc	ctttgagtga	gctgataccg	7140
	ctcgccgcag	ccgaacgacc	gagcgcagcg	agtcagttag	cgagggaagcg	gaagagcgcc	7200
65	caataccgca	accgctctc	cccgcgctt	ggccgattca	ttaatgcagc	tggcacgaca	7260
	ggtttccgga	ctggaagcg	ggcagtgagc	gcaacgcaat	taatgtgagt	tagctactc	7320
	attaggcacc	ccagcttcta	cactttatgc	ttccggctcg	tatgttgtgt	ggaattgtga	7380
	goggataaca	atttcacaca	ggaacacgct	atgaccatga	ttacgccaaag	ctggcgcg	7438

5 <210> 6
 <211> 45
 <212> DNA
 <213> Artificial Sequence

 10 <220>
 <223> Description of Artificial Sequence: primer C31-1
 <400> 6
 ataagaatgc ggccgcccga tatgacacaa ggggttgtga ccggg 45

 15 <210> 7
 <211> 40
 <212> DNA
 <213> Artificial Sequence

 20 <220>
 <223> Description of Artificial Sequence: primer C31-3
 <400> 7
 ataagaatgc ggccgcaccc gccgctacgt cttcogtgcc 40
 25

 30 <210> 8
 <211> 24
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: primer C31-8
 35 <400> 8
 cccgttggca ggaagcactt ccgg 24

 40 <210> 9
 <211> 55
 <212> DNA
 <213> Artificial Sequence

 45 <220>
 <223> Description of Artificial Sequence: primer C31-9
 <400> 9
 ggatccctcga gccgcgggcg gccgcctacg ccgctacgtc ttcogtgccg tccgtg 55
 50
 <210> 10
 <211> 5711
 <212> DNA
 <213> Artificial Sequence
 55
 <220>
 <223> Description of Artificial Sequence: vector
 pCMV-C31-Int (wt)
 60 <400> 10
 aaacagtcgc atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
 tagtaataaa ttacgggggc attagttcat agcccatata tggagttccg cgttacataa 120
 cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgccatt gacgtcaata 180
 atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
 65 tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
 cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta catgacctta 360
 tgggaacttcc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420

	cggtttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	atttccaagt	480
	ctccaccccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	540
	aaatgtcgta	acaactccgc	cccattgaog	caaatgggog	gtaggcgtgt	acgggtgggag	600
5	gtctatataa	gcagagctct	ctggctaact	agagaaccca	ctgcttactg	gcttatcgaa	660
	attaatacga	ctcactatag	ggagacccaa	gctgactcta	gacttaatta	agcgttgggg	720
	tgagtactcc	ctctcaaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	780
	cgaggaggat	ttgatattca	cctggcccgc	ggtgatgcct	ttgaggggtg	ccgcgtccat	840
	ctggtcagaa	aagacaatct	ttttgttgct	aagcttgagg	tgtggcaggc	ttgagatctg	900
	gccatacact	tgagtacat	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	960
10	cagggcgggc	gcccgatatg	acacaagggg	ttgtgaccgg	ggtggacacg	tacgggggtg	1020
	cttacgaccg	tcagtcogcg	gagcgcgaga	attcgagcgc	agcaagccca	gcgacacagc	1080
	gtagcgccaa	cgaagacaag	gcggccgacc	ttcagcgcga	agtcgagcgc	gacggggggc	1140
	ggttcagggt	cgtcgggcat	ttcagcgaag	cgcggggcac	gtcggcgctc	gggacggcgg	1200
	agcgcccga	gttcgaacgc	atcctgaacg	aatgcgcgcg	cgggcggtc	aacatgatca	1260
15	ttgtctatga	cgtgtcgcgc	ttctcgcgcg	tgaaggctcat	ggacgcgatt	cogattgtct	1320
	cggaattgct	cgccctgggc	gtgacgattg	tttccactca	ggaaggcgct	ttccggcagg	1380
	gaaacgtcat	ggacctgatt	cacctgatta	tgccgctcga	cgcgtcgcac	aaagaatctt	1440
	cgctgaagtc	ggcgaagatt	ctcgacacga	agaaccttca	gcgcgaattg	ggcgggtacg	1500
	tcggcgggaa	ggcgcccttac	ggcttcgagc	ttgtttcgga	gacgaaggag	atcacgcgca	1560
20	acggccgaat	ggccaattgtc	gtccatcaaca	agcttgcgca	ctcgaccact	cccttaccg	1620
	gaccttccga	gttcgagccc	gacgtaatcc	ggtggtggtg	gcgtgagatc	aagacgcaca	1680
	aacaccttcc	cttcaagccg	ggcagtcaag	ccgccattca	cccgggcagc	atcacggggc	1740
	tttgtaagcg	catggacgct	gacgcgcgtc	cgacccgggg	cgagacgatt	gggaagaaga	1800
	ccgcttcaag	cgccctgggac	ccggcaaccg	ttatgcgaat	ccttcgggac	ccgcgtattg	1860
25	cgggcttcgc	cgctgaggtg	atctacaaga	agaagccgga	cggcacgcgc	accacgaaga	1920
	ttgagggtta	ccgcattcag	cgcgaccoga	tcacgctccg	gccggtcgag	cttgattgcg	1980
	gaccgatcat	cgagcccgct	gagtggtatg	agcttcaggc	gtggttgagc	ggcagggggc	2040
	gcggcaaggg	gctttcccg	gggcaagcca	ttctgtccgc	catggacaag	ctgtactgcg	2100
	agtgtggcgc	cgtcatgact	tcgaagcgcg	gggaagaatc	gatcaaggac	tcttaccgct	2160
30	gccgtcgcgc	gaaggtggtc	gacccgtccg	cacctgggca	gcacgaaggc	acgtgcaacg	2220
	tcagcatggc	ggcatctgac	aagttcgttg	cgaacgcgat	cttcaacaag	atcaggcagc	2280
	ccgaaggcga	cgaagagacg	ttggcgcttc	tgtgggaagc	cgcccgacgc	ttcggcaagc	2340
	tcactgaggc	gcctgagaag	agcggcgaac	gggcgaacct	tgttgcgagg	cgcgcgcagc	2400
	ccctgaacgc	ccttgaagag	ctgtacgaag	accgcgcggc	aggcgcgtac	gacggacccg	2460
35	ttggcaggaa	gcacttccgg	aagcaacagg	cagcgcgtgac	gctccggcag	caagggggcg	2520
	aagagcggct	tgccgaactt	gaagccgcgc	aagccccgaa	gcttcccctt	gaccaatggt	2580
	tcgccgaaga	cgcgcagcgt	gacccgaccg	gccttaagtc	gtggtggggg	cgcgcgtcag	2640
	tagacgacaa	gcgcgtgttc	gtcgggctct	tcgtagacaa	gatcgttgtc	acgaagtoga	2700
	ctacgggcag	ggggcaggga	acgcccacgc	agaagcgcgc	ttcgatcacg	tgggcgaagc	2760
40	cgccgaccga	cgacgacgaa	gacgacgccc	aggacggcac	ggaagacgta	gcggcgtagg	2820
	cggcgcccgc	gctcgagatc	caggcgcgga	tcaataaaag	atcattattt	tcaatagatc	2880
	tgtgtgttgg	ttttttgtgt	gccttggggg	agggggaggc	cagaatgagg	cgcggccaag	2940
	ggggaggggg	agggcgaat	gaccttgggg	gagggggagg	ccagaatgac	cttggggggg	3000
	ggggaggcca	gaatgaggcg	cgcccccggg	taccgagctc	gaattcactg	gcggtcgttt	3060
45	tacaacgtcg	tgactgggaa	aacctggcg	ttacccaact	taatcgctt	gcagcacatc	3120
	cccttttcgc	cagctggcgt	aatagcgaag	aggcccgcac	cgatcgccct	tcccaacagt	3180
	tgcgacgctt	gaatggcgaa	tggcgccctg	tgcggtatct	tctccttacg	catctgtgcg	3240
	gtatttcaca	ccgcattatg	tgcacctoga	gtacaatctg	ctctgatgcc	gcatagttaa	3300
	gccagccccc	acaccgcgca	accccgctcg	acgcgcctcg	acgggcttgt	ctgctcccg	3360
50	catccgctta	cagacaagct	gtgaccgtct	ccgggagctg	catgtgtcag	aggttttcac	3420
	cgtcatcacc	gaaacgcgcg	agacgaaagg	gcctcgtgat	acgcctatct	ttatagggtta	3480
	atgtcatgat	aataatggtt	tcttagacgt	caggtggcac	tttccgggga	aatgtgcgcg	3540
	gaacccctat	ttgtttatct	ttctaaatac	attcaaatat	gtatccgctc	atgagacaat	3600
	aaccttgata	aatgcttcaa	taataattgaa	aaaggaagag	tatgagtatt	caacatttcc	3660
55	gtgtcgccct	tattcccttt	tttgccgcat	tttgcccttc	tgtttttgct	cacccagaaa	3720
	cgctgggtgaa	agtaaaagat	gctgaagatc	agttgggtgc	acgagtgggt	tacatcgaa	3780
	tgatctctaa	cagcggttaag	atccttgaga	gttttcgccc	cgaagaacgt	tttccaatga	3840
	tgagcacttt	taaagttctg	ctatgtggcg	cggatattatc	cogtattgac	gcggggcaag	3900
	agcaactcgg	tcgcccgcata	cactattctc	agaatgactt	ggttgagtac	tcaccagtc	3960
60	cagaaaagca	tcttacggat	ggcatgacag	taagagaatt	atgcagtgc	gccataacca	4020
	tgagtataaa	cactgcggcc	aacttacttc	tgacaacgat	cggaggaccg	aaggagctaa	4080
	ccgctttttt	gcacaacatg	ggggatcatg	taactcgcct	tgatcgttgg	gaaccggagc	4140
	tgaatgaagc	cataccaaac	gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	4200
	cgttgcgcaa	actattaaat	ggcgaactac	ttactctagc	ttcccgccaa	caattaatag	4260
65	actgattgga	ggcggataaa	gttgccaggc	cacttctgcg	ctcggccctt	ccggctggct	4320
	gttttattgc	tgataaatct	ggagccggtg	agcgtgggtc	tcgcggtatc	attgcagcac	4380
	tggggccaga	tggttaagccc	tcccgatatc	tagttatctc	cacgacgggg	agtcaggcaa	4440

6

ctatggatga acgaaataga cagatcgctg agataggtgc ctactgatt aagcattggt 4500
 aactgtcaga ccaagtttac tcatatatac tttagattga tttaaaactt catTTTTaat 4560
 ttaaaaggat ctaggtgaag atcctTTTTg ataattcat gacaaaatc ccttaaogtg 4620
 agTTTTcggt ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc 4680
 5 cTTTTTTTct ggcgtaatac tgctgcttgc aaacaaaaaa accaccgcta ccagcgggtg 4740
 tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag 4800
 cgcagatacc aaatactgtc cttctagtgt agccgtagt accgccaccac ttcagaact 4860
 ctgtagcacc gcctacatac ctgcctctgc taatcctgtt accagtggct gctgccagt 4920
 gcgataagtc gtgtcttacc ggggttggaact caagacgata gttaccggat aaggcgagc 4980
 10 ggtcgggctg aacgggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg 5040
 aactgagata cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg 5100
 cggacaggta tcgggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttocag 5160
 ggggaaacgc ctggtatctt tatagtctcg tcgggtttcg ccacctctga cttgagcgct 5220
 gattttttgt atgctcgta ggggggcgga gcctatggaa aaacgccagc aacgcggct 5280
 15 ttttacgggt cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc 5340
 ctgattctgt ggataaccgt attaccgcct ttgagtgagc tgataccgct cgcgcagcc 5400
 gaacgaccga gcgcagcgag tcagtgagcg aggaagcgga agagcgccca atacgcaaac 5460
 cgcctctccc cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg tttcccgact 5520
 ggaagcggg cagtgagcgc aacgcaatta atgtgagtta gctactcat taggcacccc 5580
 20 agcctttaca ctttatgctt ccggtcgtg tggtgtgtg aattgtgagc ggataacaat 5640
 ttcacacagg aaacagctat gaccatgatt acgccaagct agcccggtt agcttgcagt 5700
 cctgcagggt t 5711

25 <210> 11
 <211> 69
 <212> DNA
 <213> Artificial Sequence

30 <220>
 <223> Description of Artificial Sequence: primer C31-2-2

<400> 11
 35 tagaattccg ctogagagtc taaaccttcc tcttcttctt aggcgcgcgt acgttctccg 60
 tgccgtcct 69

<210> 12
 <211> 5723
 40 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: vector
 45 pCMV-C31-Int (CNLS)

<400> 12
 50 cctgcagggt taaacagtc gatgtacggg ccagatatac gcgttgacat tgattattga 60
 ctagttatta atagtaatca attacgggggt cattagttca tagcccatat atggagttcc 120
 gcgttacata acttacggta aatggccgc ctggttgacc gcccaacgac ccccgcccat 180
 tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 240
 aatgggtgga ctatttaagg taaactgcc acttggcagt acatcaagt tatcatatgc 300
 caagtacgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt 360
 acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 420
 55 ccattggtgat gcggttttgg cagtacatca atgggcgtgg atagcggttt gactcacggg 480
 gatttccaag tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac 540
 gggacttttc aaaatgtcgt aacaactccg ccccatggac gcaaatgggc ggtaggcgtg 600
 tacgggtggga ggtctatata agcagagctc tctggctaac tagagaaccc actgcttact 660
 ggcttatcga aattaatacg actcactata gggagaccca agctgactct agacttaatt 720
 60 aagcgttggg gtgagtactg cctctcaaaa gcgggcatga cttctgcgct aagattgtca 780
 gtttccaaaa acgaggagga tttgatatto ccttgcccg cggtgatgcc tttgagggtg 840
 gcgcgctcca tctggtcaga aaagacaatc tttttgttgt caagcttgag gtgtggcagg 900
 cttgagatct ggccatacac ttgagtgaac ttgacatcca ctttgccttt ctctccacag 960
 gtgtccactc ccaggggcgc cgcgcgatat gacacaaggg gttgtgaccg ggggtggcac 1020
 65 gtacgcgggt gcttacgacc gtcagtgcgc cgagcgcgag aattcgagcg cagcaagccc 1080
 agcgacacag cgtagcgcca acgaagaca ggccggccg cttcagcgcg aagtcgagcg 1140
 cgacgggggc cggttcagggt tcgtcgggca tttcagcgaa gcgcgggca cgtcggcgtt 1200

	cgggacggcg	gagcgcccg	agttcgaa	catctgaac	gaatgccg	ccgggcggt	1260
	caacatgate	attgtctatg	acgtgtcg	cttctcgcg	ctgaaggta	tggacgcat	1320
	tccgattgtc	tccgaattgc	tccgctggg	cgtgacgatt	gtttccactc	aggaaggcgt	1380
5	cttccggcag	ggaaacgtca	tggacctgat	tcacctgatt	atgcggtctg	acgcgtcgca	1440
	caaagaatct	tgcgtgaagt	cggcgaagat	tctcgacacg	aagaaccttc	agcgcggaatt	1500
	gggcggttac	gtcggcgga	aggcgctta	cggcttcgag	cttgtttcgg	agacgaagga	1560
	gatcacgcgc	aacggccgaa	tggtaaatgt	cgtcatcaac	aagcttgccg	actcgaccac	1620
	tcccttacc	ggaccttctg	agttcgagcc	cgacgtaatc	cgggtggtgt	ggcgtgagat	1680
10	caagacgcac	aaacaccttc	ccttcaagcc	gggcagtcaa	gccgccatto	acccgggcag	1740
	catcacgggg	ctttgttaagc	gcattggacgc	tgacgccgtg	ccgaccgggg	gcgagacgat	1800
	tgggaagaag	accgcttcaa	gcgcctggga	cccggcaacc	gttatgcgaa	tccttcggga	1860
	ccgcgctatt	gcgggcttgc	ccgctgaggt	gatctacaag	aagaagccgg	acggcacgccc	1920
	gaccacgaag	attgagggtt	accgcattca	cgcgaccggg	atcacgctcc	ggcgggtcga	1980
15	gcttgattgc	ggacogatac	tccgacccgc	tgagtggat	gagcttcagg	cgtggttga	2040
	cgccaggggg	cgccgcaagg	ggctttcccg	ggggcaagcc	attctgtccg	ccatggacaa	2100
	gctgtactgc	gagtgtggcg	ccgtcatgac	ttcgaagcgc	ggggaagaat	cgatcaagga	2160
	ctcttaccgc	tgcgctcgcc	ggaaggtggt	cgaccctgcc	gcacctgggc	agcacgaagg	2220
	cacgtgcaac	gtcagcatgg	cggcactcga	caagttcggt	gcggaacgca	tcctcaacaa	2280
20	gatcaggcac	ccgaagggcg	acgaagagac	gttggcgctt	ctgtgggaag	ccgcccgagc	2340
	cttcggcaag	ctcactgagg	cgccctgagaa	gagcgccgaa	cgggcggaacc	ttgttcggga	2400
	gcgcgcccgc	gccctgaacg	cccttgaaga	gctgtacgaa	gaccgcgcgg	caggcgcgta	2460
	cgacggaccc	gttggcagga	agcacttcgg	gaagcaacag	gcagcgctga	cgctccggca	2520
	gcaagggggc	gaagagcggc	ttgccgaact	tgaagccgcc	gaagcccgga	agcttccccc	2580
25	tgaccaattg	ttcccggaag	acggccgagc	tgaccgcgac	ggccctaagt	cgtggtgggg	2640
	gcgcgcgtca	gttagacgaca	agcgcgtggt	cgtcggtgct	ttcgtagaca	agatcgttgt	2700
	cacgaagtgc	actacgggca	ggggcgagg	aacgcccac	gagaagcgcg	cttcgatcac	2760
	gtgggcgaag	ccgcgcgacc	acgacgacga	agacgacgcc	caggacggca	cggaagacgt	2820
	agcggcgcc	aagaagaaga	ggaaggttta	gactctcgag	atccaggcgc	ggatcaataa	2880
30	aagatcatta	ttttcaatag	atctgtgtgt	tggttttttg	tgtgccttgg	gggaggggga	2940
	ggccagagag	aggcgcgccc	aagggggagg	gggaggccag	aatgaccttg	ggggaggggg	3000
	aggccagaat	gaccttgggg	ccagaatgag	ccagaatgag	gcgcgcccc	gggtaccgag	3060
	ctcgaattca	ctggcgctcg	ttttacaacg	tctgtaactg	gaaaaccttg	gcgttaccga	3120
	acttaatcgc	cttgacgac	atcccccttt	cgccagctgg	cgtaatagcg	aagaggcccg	3180
35	caccgatcgc	ccttcccaac	agttgcccag	cctgaatggc	gaatggcgcc	tgatgcggta	3240
	ttttctcctt	acgcatactg	gcggtatttc	acacgcgata	tgggtgactc	tcagtacaat	3300
	ctgctctgat	ccgcgcatag	taagccagcc	ccgacacccg	ccaacacccg	ctgacgcgcc	3360
	ctgaagggtc	tgtctgctcc	cgccatccgc	ttacagacaa	gctgtgaccg	tctccgggag	3420
	ctgcatgtgt	cagaggtttt	cacgcgtcat	accgaaacgc	gcgagacgaa	agggcctcgt	3480
40	gatacgccta	tttttatagg	ttaatgtcat	gataataatg	gtttcttaga	cgtcagggtg	3540
	cacttttcgg	ggaaatgtgc	gcggaaacccc	tatttgttta	tttttctaaa	tacattcaaa	3600
	tatgtatccg	ctcatagacg	aataacccgt	ataaatgctt	caataatatt	gaaaaaggaa	3660
	gagtatgagt	attcaacatt	tccgtgtcgc	ctctattccc	ttttttgcgg	cattttgcct	3720
	tcctgttttt	gctcaccag	aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	3780
45	tgcacgagtg	ggttacatcg	aactggatct	caacagcgg	aagatccctg	agagttttcg	3840
	cccggaagaa	cgttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	gcgcgggtatt	3900
	atcccgctatt	gacgcggggc	aagagcaact	cggctcgccg	atacactatt	ctcagaatga	3960
	cttgggttag	tactcaccag	tcacagaaaa	gctacttag	gatggcatga	cagtaagaga	4020
	attatgcagt	gctgccataa	ccatgagtga	taacactcgc	gccaaacttac	ttctggaac	4080
50	gatcggagga	ccgaaggagc	taaccgcttt	tttgcaaac	atgggggatc	atgtaactcg	4140
	ccttgatcgt	tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	gtgacaccac	4200
	gatgcctgta	gcaatggcaa	caaogttgog	caaactatta	actggcgaa	tacttactct	4260
	agcttcccg	caacaattaa	tagactggat	ggaggcggat	aaagttgcag	gaccacttct	4320
	gcgctcggcg	cttcggctg	gctgggttat	tgctgataaa	tctggagccg	gtgagcgtgg	4380
55	gtctcgggt	atcattgcag	cactggggcc	agatggtaag	ccctcccgta	tcgtagtatt	4440
	ctacacgacg	gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	4500
	tgctcactg	attaagcatt	ggtaactgtc	agaccaagtt	tactcatata	tacttttagat	4560
	tgatttaaaa	cttcattttt	aatttaaaag	gatctaggtg	aagatccctt	ttgataatct	4620
	catgaccaa	atcccttaac	gtgagttttt	gttccactga	gcgtcagacc	ccgtagaaaa	4680
60	gatcaaaagg	ttctcttgag	atcctttttt	tctgcgogta	atctgctgct	tgcaaacaaa	4740
	aaaaccaccc	ctaccagcgg	tggtttgttt	gcgggatcaa	gagctaccaa	ctctttttcc	4800
	gaaggtaact	ggcttcagca	gagcgcagat	accaaatact	gtcctttctag	tgtagccgta	4860
	gttaggccc	cacttcaaga	actctgtagc	accgcctaca	tacctcgctc	tgctaactct	4920
	gttaccagtg	gctgctgcca	gtggcgataa	gtcgtgtott	accgggttgg	actcaagacg	4980
65	atagttaccg	gataaggcgc	agcggctggg	ctgaacgggg	ggttcgtgca	cacagcccag	5040
	cttggagcga	acgaactgag	ccgaactgag	atacctacag	cgtgagctat	gagaaagcgc	5100
	cacgcttccc	gaaggagaa	aggcggacag	gtatccggta	agcggcaggg	tcgggaacgg	5160
	agagcgcacg	aggagccttc	cagggggaaa	cgccctggat	ctttatagtc	ctgtcgggtt	5220

	togccacctc	tgacttgagc	gtcgattttt	gtgatgctcg	tcaggggggc	ggagcctatg	5280
	gaaaaacgcc	agcaacgcgg	ccttttttacg	gttcctggcc	ttttgctggc	cctttgctca	5340
	catgttcttt	cctgcgttat	cccctgatcc	tgtggataac	cgtattaccg	cctttgagtg	5400
5	agctgatacc	gctcgccgca	gccgaacgac	cagagcgagc	gagtcagtga	gcgaggaagc	5460
	ggaagagcgc	ccaatacgca	aaccgcctct	ccccgcgct	tggccgattc	attaatgcag	5520
	ctggcacgac	aggtttccc	actggaaagc	gggcagtgag	cgcaacgcaa	ttaatgtgag	5580
	ttagctcaact	cattaggcac	cccaggcttt	acactttatg	cttcgggctc	gtatgttggtg	5640
	tgggaattgtg	agcggataac	aatttcacac	aggaaacagc	tatgaccatg	attacgccaa	5700
10	gctagcccg	gctagcttgc	atg				5723

<210> 13

<211> 4960

<212> DNA

15 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pCMV-Cre

20

<400> 13

	aaacagtccg	atgtacgggc	cagatatacg	cgttgacatt	gattattgac	tagttattaa	60
	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg	cgttacataa	120
25	ccttacggtaa	atggcccggc	tggctgaccc	cccaacgacc	cccgccatt	gacgtcaata	180
	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	atgggtggac	240
	tatttacggg	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	aagtagcccc	300
	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	catgacctta	360
	tgggactttc	ctacttggca	gtacatctac	gtattagtc	tcgctattac	catggtgatg	420
30	cgggttttgg	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	atttccaagt	480
	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	540
	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	acggtgggag	600
	gtctatataa	gcagagctct	ctggctaact	agagaaccca	ctgcttactg	gcttatcgaa	660
	attaatacga	ctcactatag	ggagacccaa	gctgactcta	gacttaatta	agcgttgggg	720
	tgagtactcc	ctctcaaaa	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	780
35	cgaggaggat	ttgatattca	cctggcccgc	ggtgatgcct	ttgaggggtg	ccgcgtccat	840
	ctggtcagaa	aagacaatct	ttttgttgct	aagcttgagg	tgtggcaggc	ttgagatctg	900
	gccatacact	tgagtgcac	tgacatccac	tttgcctttc	totccacagg	tgtccactcc	960
	cagggcggcc	tcgaccatgc	ccaagaagaa	gaggaagggt	tccaatttac	tgaccgtaca	1020
40	ccaaaatttg	cctgcattac	cggctcgatg	aacgagtgat	gaggttcgca	agaacctgat	1080
	ggacatgttc	agggatcgcc	aggcgttttc	tgagcatacc	tggaaaatgc	ttctgtccgt	1140
	ttgcggttgc	tgggcggcat	ggtgcaagtt	gaataacccg	aaatggtttc	ccgcagaacc	1200
	tgaagatgtt	cgcgattatc	ttctatatct	tcaggcgcgc	ggtctggcag	taaaaactat	1260
	ccagcaacat	ttggggccag	ttaacatgct	tcctcgtcgg	tcggggctgc	cacgaccaag	1320
45	tgacagcaat	gctgtttcac	tggttatgcg	gcggatccga	aaagaaaacg	ttgatgcggg	1380
	tgaacgtgca	aaacaggctc	tagcgttcga	acgcactgat	ttcgaccagg	ttcgttcact	1440
	catggaaaa	agcgtatcgt	gccaggatat	agctaactcg	gcattttctg	ggattgtcta	1500
	taacaccctg	ttacgtatag	ccgaatttgc	caggatcagg	gttaaagata	tctcacgtac	1560
	tgacggtggg	agaatgttaa	tccatatttg	cagaacgaaa	acgtgtggtt	gcaccgcagg	1620
50	tgtagagaag	gcacttagcc	tgggggtaac	taaactggtc	gagcgatgga	tttccgtctc	1680
	tgggtgtagc	gatgatccga	ataactacct	gttttgccgg	gtcagaaaaa	atggtgttgc	1740
	cgcgccatct	gccaccagcc	agctatcaac	tcgcgcctcg	gaagggtatt	ttgaagcaac	1800
	tcctcgattg	atttacggcg	ctaaggatga	ctctggtcag	agatacctgg	cctggtctgg	1860
	acacagtgcc	cgtgtcggag	ccgcgcgaga	tatggcccg	gctggagttt	caataccgga	1920
55	gatcatgcaa	gctggtggct	ggaccaatgt	aaatattgtc	atgaactata	tccgtaacct	1980
	ggatagtga	acaggggcaa	tgggtgcgct	gctggaagat	ggcgattagc	cattaacgcg	2040
	taaatgattg	cagatccact	agttctaggg	ccgcgtcgac	ctcgagatcc	aggcgcggat	2100
	caataaaa	tcattatttt	caatagatct	gtgtgttgg	tttttgtgtg	ccttggggga	2160
	gggggaggcc	agaatgaggc	gcggccaagg	gggaggggga	ggccagaatg	accttggggg	2220
	agggggaggc	cagaatgacc	ttgggggagg	gggagggcag	aatgaggcgc	gccccgggt	2280
60	accgagctcg	aattcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accttggcgt	2340
	tacccaactt	aatcgcttgc	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	2400
	ggcccgcaac	gatcgccctt	cccaacagtt	gcgcagcctg	aatggcgaat	ggcgctgat	2460
	gcggtatttt	ctccttacgc	atctgtgcgg	tatttcacac	cgcatatggt	gcactctcag	2520
	tacaatctgc	tctggtcccg	catagttaa	ccagcccgca	caccgcgcaa	caccgcgta	2580
65	cgcgccttga	cgggcttgtc	tgctcccggc	atccgcttac	agacaagctg	tgaccgtctc	2640
	cgggagctgc	atgtgtcaga	ggttttcacc	gtcatcaccg	aaacgcgcga	gacgaaagg	2700
	cctcgtgata	cgccatattt	tatagggttaa	tgtcatgata	ataatgggtt	ccttagacgtc	2760

	aggtggcact	tttcggggaa	atgtgcgcgg	aaccocctatt	tgttttatttt	tctaaatata	2820
	ttcaaatatg	tatccgctca	tgagacaata	aoccttgataa	atgcttcaat	aatattgaaa	2880
	aaggaagagt	atgagtattc	aacattttcog	tgtgcgccctt	attccctttt	ttgcggcatt	2940
5	ttgccttcct	gtttttgtct	accagaaac	gctgggtgaaa	gtaaaagatg	ctgaagatca	3000
	gttgggtgca	cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	3060
	ttttcgcccc	gaagaacgtt	ttccaatgat	gagcaactttt	aaagtctctc	tatgtggcgc	3120
	ggattatcc	cgtattgaog	ccgggcaaga	gcaactcggg	cgcgcatac	actattctca	3180
	gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	3240
	aagagaatta	tgcagtgtcg	ccataaccat	gagtataaac	actgcggcca	acttacttct	3300
10	gacaacgatc	ggaggaccga	aggagctaac	cgctttttttg	cacaacatgg	gggatcatgt	3360
	aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaaacg	acgagcgtga	3420
	caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	3480
	tactctagct	tcccggaac	aattaataga	ctggatggag	gcggataaag	ttgcaggacc	3540
	acttctgcgc	tccggcccttc	cggttggtcg	gtttattgct	gataaatctg	gagcgggtga	3600
15	gcgtgggtct	cgcggtatca	ttgcagcaact	ggggccagat	ggtaagccct	cccgtatcgt	3660
	agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgtgta	3720
	gatagggtgc	tacttgatta	agcatttggt	actgtcagac	caagtttact	catatatact	3780
	ttagattgat	ttaaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	tccttttttga	3840
	taatctcatg	accaaaatcc	cttaacgtga	gttttctgttc	cactgagcgt	cagaccccg	3900
20	agaaaagatc	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaactc	gctgcttgca	3960
	aacaaaaaaa	ccaccgctac	cagcgggtgg	ttgtttgccc	gatcaagagc	taccaactct	4020
	ttttccgaag	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgt	4080
	gccgtagtta	ggccaccact	tcaagaaactc	tgtagcacccg	cctacatacc	tcgctctgct	4140
	aatcctgtta	ccagtggctg	ctgccagtg	cgataagtcg	tgtcttaccg	ggttggaactc	4200
25	aagacgatag	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggtt	cgtgcacaca	4260
	gccagccttg	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	agctatgaga	4320
	aagcgcacag	cttcccgaag	ggagaaaggc	ggacaggtat	ccggttaagc	gcagggtcgg	4380
	aacaggagag	cgcacgaggg	agcttccagg	gggaaacgcc	tggatctctt	atagtcctgt	4440
	cgggtttcgc	cacctctgac	ttgagcgtcg	atttttgtga	tgtcgtcag	gggggcggag	4500
30	cctatgaaa	aacgcacga	acgcggcctt	tttaacggtt	ctggcctttt	gctggccttt	4560
	tgtcctcatg	ttctttcctg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	4620
	tgagtgaact	gataccgctc	gdcgcagccg	aacgacccag	cgcagcaggt	cagtgaagca	4680
	ggaagcggaa	gagcgcacca	taocgcaaac	gcctctcccc	gcgcgttggc	cgattcatta	4740
	atgcagctgg	cacgacaggt	ttcccgactg	gaaagcgggc	agtgaagcga	acgcaattaa	4800
35	tgtgagttag	ctcaactcatt	aggcaccoca	ggcttttacac	tttatgcttc	cggctcgtat	4860
	ggtgtgtgga	attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	4920
	cgccaagcta	gcccgggcta	gcttgcatgc	ctgcagggtt			4960
40	<210> 14						
	<211> 3858						
	<212> DNA						
	<213> Artificial Sequence						
45	<220>						
	<223> Description of Artificial Sequence: vector pRK50						
	<400> 14						
50	aaacagtcog	atgtacgggc	cagatatacg	cgttgacatt	gattattgac	tagttattaa	60
	tagtaatcaa	ttacgggggc	attagttcat	agcccatata	tggagttccg	cgttacataa	120
	cttacggtaa	atggcccgcc	tggttgaccg	cccaacgacc	cccgccatt	gacgtcaata	180
	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	atgggtggac	240
	tatttacggg	aaactgcccc	cttggcagta	catcaagtg	atcatatgcc	aagtacgccc	300
	cctattgacg	tcaatgacgg	taaatggccc	gcttggcatt	atgccagta	catgacctta	360
55	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	catgggtgatg	420
	cggtttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	atttccaagt	480
	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	540
	aaatgtcgt	acaactccgc	ccatttgacg	caaatggggc	gtaggcgtgt	acgggtgggag	600
	gtctatataa	gcagagctct	ctggctaact	agagaaacca	ctgcttactg	gcttatcgaa	660
60	attaatacga	ctcactatag	ggagaoccaa	gctgactcta	gacttaatta	agcgttgggg	720
	tgagtactcc	ctctcaaaa	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	780
	cgaggaggat	ttgatattca	cctggcccg	ggtgatgcct	ttgagggtgg	ccgcgtccat	840
	ctggtcagaa	aagacaatct	ttttgttgct	aagcttgagg	tgtggcaggc	ttgagatctg	900
	gccatacact	tgagtgcact	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	960
65	cagggcggcc	gcgtgcacct	cgagatccag	gcgcggatca	ataaaagatc	attattttca	1020
	atagatctgt	gtgttggttt	tttgtgtg	ttggggagg	gggaggcca	aatgaggcgc	1080
	ggccaagggg	gagggggagg	ccagaatgac	cttggggagg	ggggaggcca	gaatgacctt	1140

	gggggagggg	gaggccagaa	tgaggcgogc	ccccgggtac	cgagctcgaa	ttcactggcc	1200
	gtcgtttttac	aacgtcgtga	ctgggaaaac	cctggcggtta	cccaacttaa	tcgccttgca	1260
	gcacatcccc	ctttcgccag	ctggcgtaat	agcgaagagg	cccgcaccga	tcgccccttc	1320
	caacagttgc	gcagcctgaa	tggcgaaatg	cgctgatgc	ggtattttct	ccttacgcac	1380
5	ctgtgcggta	tttcacaccg	catatgggtg	actctcagta	caatctgctc	tgatgccgca	1440
	tagttaagcc	agccccgaca	cccgccaaga	cccgctgacg	cgccctgacg	ggcttgtctg	1500
	ctcccggcat	ccgcttacag	acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	1560
	ttttcaccgt	catcaccgaa	acgcgcgaga	cgaaagggcc	tcgtgatagc	cctattttta	1620
	taggttaaatg	tcattgataat	aatggttttc	tagacgtcag	gtggcaacttt	tcggggaaat	1680
10	gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaatatgta	tcgcctcatg	1740
	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa	1800
	catttccgtg	tcgccccttat	tccttttttt	gcggcatttt	gccttcctgt	ttttgctcac	1860
	ccagaaacgc	tgggtgaaagt	aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac	1920
	atcgaaactgg	atctcaacag	cggtaagatc	ccttgagagtt	ttcgccccga	agaacgtttt	1980
15	ccaatgatga	gcactttttaa	agttctgcta	tgtggcgogg	tattatcccg	tattgacgcc	2040
	gggcaagagc	aactcggctg	ccgcatacac	tattctcaga	atgacttggt	tgagtactca	2100
	ccagtcacag	aaaagcatct	tacggatggc	atgacagtaa	gagaattatg	cagtgtctgc	2160
	ataaccatga	gtgataaac	tgccggccaa	ttactttotga	caacgatcgg	aggaccgaag	2220
	gagctaaccg	ctttttttgca	caacatgggg	gatcatgtaa	ctcgccctga	tcgttgggaa	2280
20	ccggagctga	atgaagccat	accaaagcac	gagcgtgaca	ccacgatgcc	tgtagcaatg	2340
	gcaacaacgt	tgcgcaaaact	attaactggc	gaactactta	ctctagcttc	ccggcaacaa	2400
	ttaatagact	ggatggaggg	ggtataaagt	gcaggaccac	ttctgcgctc	ggcccttcgc	2460
	gctggctggt	ttattgctga	taaatctgga	gcccgtgagc	gtgggtctcg	cggtatcatt	2520
	gcagcactgg	ggccagatgg	taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	2580
25	caggcaacta	tggatgaacg	aaatagacag	atcgtctgaga	taggtgcctc	actgattaag	2640
	cattgggtaac	tgtcagacca	agtttaactca	tatatacttt	agattgattt	aaaacttcac	2700
	ttttaatttta	aaaggatcta	ggtgaagatc	cttttttgata	atctcatgac	caaaatccct	2760
	taacgtgagt	tttcggttcca	ctgagcgta	gaccccgtag	aaaagatcaa	aggatcttct	2820
	tgagatcctt	tttttctgcg	cgtaatctgc	tgcttgcaaa	caaaaaaacc	accgctacca	2880
30	gcggtggttt	gtttgccgga	tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	2940
	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	3000
	aagaaactctg	tagcaccgcc	tacatacctc	gctctgctaa	tcctgttacc	agtggctgct	3060
	gccagtggcg	ataagtcgtg	tcttacccgg	ttggactcaa	gacgatagtt	accggataag	3120
	gcgcagcgg	cgggctgaac	gggggggttc	tgacacacag	ccagcttgga	gcgaacgacc	3180
35	tacaccgaac	tgagatacct	acagcgtgag	ctatgagaaa	gcgccacgct	tcgccgaagg	3240
	agaaaggcgg	acaggtatcc	ggtaagcggc	agggtcgga	caggagagcg	cacgagggag	3300
	cttccagggg	gaaacgcctg	gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	3360
	gagcgtcgat	ttttgtgatg	ctcgctcagg	ggcgaggacc	tatggaaaaa	cgccagcaac	3420
	gcggcctttt	tacggttcct	ggcctttttg	tgcccttttg	ctcacatggt	ctttcctgcg	3480
40	ttatcccctg	attctgtgga	taaccgtatt	accgcctttg	agtgagctga	taccgctcgc	3540
	cgcagccgaa	cgaccgagcg	cagcgagtoa	gtgagcgagg	aagcgggaaga	gcgcccaata	3600
	cgcaaacccg	ctctcccgcg	gcgttggccg	attcattaat	gcagctggca	cgacaggttt	3660
	cccgaactga	aagcgggcag	tgagcgcaac	gcaattaatg	tgagttagct	cactcattag	3720
	gcaccccag	ctttacactt	tatgottccg	gctcgtatgt	tgtgtggaat	tgtgagcgga	3780
45	taacaatttc	acacaggaaa	cagctatgac	catgattacg	ccaagctagc	ccgggctagc	3840
	ttgcatgoc	gcaggttt					3858

<210> 15

<211> 6257

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pRK64(deltaCre)

<400> 15

	cgctcatcac	gaaacgcgcg	aggcagctgt	ggaatgtgtg	tcagttaggg	tgtggaaagt	60
60	cccaggctc	cccagcaggc	agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	120
	ggctccccag	caggcagaag	tgtgcaaagc	atgcatctca	attagtcagc	aaccatagtc	180
	ccgcccctaa	ctccgcccac	cccgccccta	actccgcccc	gttccgcccc	ttctccgccc	240
	catggctgac	taattttttt	tatttatgca	gaggccgagg	ccgcctcggc	ctaggaacag	300
	tcgacgacac	tgacagagac	tacttcaacta	acaaccggta	cagttcgtgg	accagatggg	360
65	tgagggtggg	tacgcgcccg	gggagcccaa	gggcacggcc	tggcacccgc	accgcggctt	420
	cgagaccgtc	acgaatagat	ccataacttc	gtatagcata	cattatacga	agttataccg	480
	ggccaccatg	gtcgcgagta	gcttggcaact	ggccgtcggt	ttacaaacgtc	gtgactggga	540

	aaaccctggc	gttacccaac	ttaatcgcc	tgcagacac	cccccttgc	ccagctggcg	600
	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	660
	atggcgcttt	gcctgggttc	cggaaccaga	agcgggtgcc	gaaagctggc	tggagtgcga	720
	tcttcctgag	gcgatactg	tcgctgtccc	ctcaaactgg	cagatgcacg	gttacgatgc	780
5	gcccattctac	accaacgtaa	cctatcccat	tacgggtcaat	cgcgcgtttg	ttcccacgga	840
	gaatccgacg	ggttggttact	cgctcacatt	taatgttgat	gaaagctggc	tacaggaagg	900
	ccagacgcga	attatttttg	atggcggttaa	ctcggcgttt	catctgtggg	gcaacggggc	960
	ctgggtcggt	tacggccagg	acagtcgttt	gocgtctgaa	tttgacctga	gcgcattttt	1020
10	acgcgcggga	gaaaaccgcc	tcgoggtgat	ggtgctgcgt	tggagtgcag	gcagttatct	1080
	ggaagatcag	gatattgtgg	ggatgagcgg	cattttccgt	gacgtctcgt	tgctgcataa	1140
	accgactaca	caaatacagc	atttccatgt	tgccactcgc	tttaattgat	atttcagccg	1200
	cgctgtactg	gaggctgaag	ttcagatgtg	cggcgagttg	cgtgactacc	tacgggtaac	1260
	agtttcttta	tggcagggtg	aaacgcaggt	cgccagcgcc	accgcgcctt	tcggcggtga	1320
	aattatcgat	gagcgtggtg	gttatgcgga	tcgcttcaca	ctacgtctga	acgtcgaaaa	1380
15	cccgaaactg	tggagcgccg	aaatcccgaa	tctctatcgt	gcggtggttg	aactgcacac	1440
	cgccgacggc	acgctgattg	aagcagaagc	ctgcgatgtc	ggtttccgcg	aggtgcggat	1500
	tgaaaatggt	ctgctgctgc	tgaacggcaa	gocgttgctg	attcgaggcg	ttaacctgtc	1560
	cgagcatcat	cctctgcatt	gtcaggctcat	ggatgagcag	acgatggtgc	aggatatcct	1620
20	gctgatgaag	cagaacaact	ttaacgcctt	gcatgtgttc	cattatccga	accatccgct	1680
	gtggtacacg	ctgtgcgacc	gctacggcct	gtgtgtggtg	gatgaagcca	atattgaaac	1740
	ccacggcatg	gtgccaatga	atcgtctgac	cgatgatccg	cgctggctac	cgccgatgag	1800
	cgaacgcgta	acgcgaatgg	tgcagcgoga	tcgtaatac	ccgagtgtga	tcattctggtc	1860
	gctggggaat	gaatcaggcc	acggcgctaa	tcacgacgcg	ctgtatcgct	ggatcaaato	1920
25	tgtcgatcct	tccgcgcccg	tgcagtatga	agggcgcgga	gccgacacca	cgccaccgga	1980
	tattattttg	ccgatgtacg	cgogctgtga	tgaagaccag	cccttcccg	ctgtgcgaaa	2040
	atggtccatc	aaaaaatggc	tttcgctacc	tggagagacg	cgcccgctga	tcctttgcga	2100
	atacggccac	gcgatgggta	acagtccttg	cggtttcgct	aaatactggc	agggctttcg	2160
	tcagtatccc	cgtttacagg	gcggcttcgt	ctgggactgg	gtggatcagt	cgctgattaa	2220
	atatgatgaa	aacggcaacc	cgtggtcggc	ttacggcggt	gattttggcg	atacggcgaa	2280
30	cgatcgccag	ttctgtatga	acggtctggt	ctttgcccgc	cgacgcgcgc	atccagcgct	2340
	gacggaagca	aaacaccagg	agcagttttt	ccagttccgt	ttatccgggc	aaaccatcga	2400
	agtaccagc	gaatacctgt	tcgctcatag	cgataacgag	ctcctgcact	ggatggtggc	2460
	gctggatggt	aagccgctgg	caagcggtga	agtgcctctg	gatgtcgctc	cacaaggtaa	2520
	acagttgatt	gaactgcctg	aactaccgca	gcccggagagc	gcccggcaac	tctggctcac	2580
35	agtacgcgta	gtgcaaccga	acgcgaaccg	atggtcagaa	gcccggcaca	tcagcgccctg	2640
	gcagcagtg	cgtctggcgg	aaaacctcag	tgtgacgctc	cccgcgcgct	cccacgccat	2700
	ccgcgactcg	accaccagcg	aaatggattt	ttgcatacgag	ctgggttaata	agcgttgcca	2760
	atttaaccgc	cagtcaggct	ttctttcaca	gatgtggatt	ggcgataaaa	aacaactgct	2820
	gacgcgcgct	cgcgatcagt	tcacccgtgc	accgctggat	aacgacattg	gcgtaagtga	2880
40	agcgaccgcg	attgacccta	acgcctgggt	cgaacgctgg	aaggcgggcg	gccattacca	2940
	ggccgaagca	gcgttggtgc	agtgcacggc	agatacactt	gctgatgcgg	tgctgattac	3000
	gaccgcctcac	gcgtgcagc	aaccttatct	atcagccgga	atcagccgga	aaacctaccg	3060
	gattgatggt	agtggtcaaa	tggcgattac	cggttgatgtt	gaagtggcga	gcgataaccc	3120
	gcatccggcg	cggattggcc	tgaactgcca	gctggcgagc	gtagcagagc	gggtaaactg	3180
45	gctcggatta	gggcccgaag	aaaactatoc	cgaccgcctt	actgcccgcct	gttttgaccg	3240
	ctgggatctg	ccattgtcag	acatgtatac	cccgtacgtc	ttcccagagc	aaaacggtct	3300
	gcgctgcggg	acgcgcgaat	tgaatttatg	cccacaccag	tggcgcgggc	acttccagtt	3360
	caacatcagc	cgcatacagtc	aacagcaact	cggtgaaacc	agccatcgcc	atctgtcgga	3420
	cgcggaagaa	ggcacatggc	tgaatatcga	cgggtttccat	atgggggattg	gtggcgacga	3480
50	ctcctggagc	cgcctcagat	cggcggaatt	ccagctgagc	gcccgtcgct	accattacca	3540
	gttggctctg	tgtcaaaaat	aataataacc	gggcaggggg	gatctttgtg	aaggaaacctt	3600
	acttctgttg	tgtgacataa	ttggacaaac	tacctacaga	gatttaaagc	tctaaggtaa	3660
	atataaaatt	tttaagtgtg	taatgtgtta	aactactgat	tctaattggt	tgtgtatttt	3720
	agattccaac	ctatggaact	gatgaatggg	agcagtggtg	gaatgccaga	tccagacatg	3780
55	ataagataca	ttgatgagtt	tggacaaaac	acaactagaa	tgcagtgaag	aaaatgcttt	3840
	atttgtgaaa	tttgtgatgc	tattgtctta	tttgtaacca	ttataagctg	caataaacia	3900
	gttaacaaca	acaattgcat	tcatttttatg	tttcagggttc	agggggaggt	gtgggaggtt	3960
	ttttaagca	agtaaaacct	ctacaaatgt	ggtatggctg	attatgatct	gcggccgcag	4020
	ggcctcgtga	taagcctatt	tttatagggt	aatgtcctga	taataatggt	ttcttagacg	4080
60	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	tttctaataa	4140
	cattcaaata	tgtatccgct	catgagacaa	taacctgat	aaatgcttca	ataatattga	4200
	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttgcggca	4260
	ttttgccttc	ctgtttttgc	tcacccagaa	acgctgggtg	aagtaaaaga	tgctgaagat	4320
	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	4380
65	agttttcgcc	ccgaagaacg	ttttccaatt	atgagcactt	ttaaagttct	gctatgtggc	4440
	gcggtattat	cccgatttga	cgcggggcaa	gagcaactcg	gtcgccgcat	acactattct	4500
	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggcatgaca	4560

gtaagagaat tatgcagtgc tgcataaacc atgagtata acactgcggc caacttactt 4620
 ctgacaacga tgcgaggacc gaaggagcta accgcttttt tgcacaacat gggggatcat 4680
 gtaactcgcc ttgatcggtt ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt 4740
 5 gacaccacga tgcctgtagc aatggcaaca acgttgcgca aactattaac tggcgaacta 4800
 cttactctag cttcccgcca acaattaata gactggatgg aggcggataa agttgcagga 4860
 ccacttctgc gctcgccctt tccggctggc tggtttattg ctgataaatac tggagccggg 4920
 gagcgtgggt ctgcgggtat cattgcagca ctggggccag atggtaagcc ctcccgatc 4980
 gtagttatct acacgacggg gagtacggca actatggatg aacgaaatag acagatcgct 5040
 10 gagataggtg cctcaactgat taagcattgg taactgtcag accaagttta ctcatatata 5100
 ctttagattg atttaaaact tcaatttttaa tttaaaagga tctaggtgaa gatccttttt 5160
 gataatctca tgacccaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc 5220
 gtagaaaaga tcaaaaggatc ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg 5280
 caaacaaaaa aaccaccgct accagcgggt gtttggttgc cggatcaaga gctaccaact 5340
 15 ctttttccga aggttaactgg cttcagcaga gcgcagatac caaatactgt ccttctagt 5400
 tagccgtagt taggcaccca cttcaagaac tctgtagcac cgcctacata cctcgtctg 5460
 ctaatcctgt taccagtggc tgcgtccagt ggcgataagt cgtgtcttac cgggttggac 5520
 tcaagacgat agttaccgga taaggcgcag cggctcgggt gaacgggggg ttcgtgcaca 5580
 cagcccagct tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga 5640
 20 gaaagcgcca cgcttccoga agggagaaag gcggacaggt atccggtaag cggcagggc 5700
 ggaacaggag agcgacagag ggagcttcca gggggaaacg cctggtatct ttatagtct 5760
 gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctctc agggggggcg 5820
 agcctatgga aaaacgccag caacgcggcc tttttacggt tccctggcct ttgctggcct 5880
 tttgctcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc 5940
 25 tttgagttag ctgataccgc tgcgcgcagc cgaacgaccg agcgcagcga gtcagtgage 6000
 gaggaagcgg aagagcgccc aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat 6060
 taatgcagct ggcacgcagc gtttcccgac tggaaagcgg gcagtgcgc caacgcaatt 6120
 aatgtgagtt agctcaactc ttaggcaccc caggctttac actttatgct tccggctcgt 6180
 atgttgtgtg gaattgtgag cggataacaa tttcacacag gaaacagcta tgaccatgat 6240
 30 tacgccaagc tggcgcg 6257

<210> 16

<211> 6252

<212> DNA

35 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pRK64 (deltaInt)

40

<400> 16

cgtcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagttaggg tgtggaaagt 60
 cccagggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag tcagcaacca 120
 45 ggctccccag caggcagaag tgtgcaaagc atgcatctca attagtcagc aaccatagtc 180
 ccgcccctaa ctccgcccat ccgcgcccta actccgcca gttccgcca ttctccgcc 240
 catggttgac taattttttt tatttatgca gaggccgagg ccgcctcggc ctaggaaacg 300
 togacgacac tgcagagacc tacttcaact acaaccggta cagttcgtgg accagatggg 360
 tgaggtggag tacgcgccg gggagcccaa aggttacccc agttggggca ctactccga 420
 50 aaaccgcttc tggatccata acttcgtata gcatacatia tacgaagtta taccggggcca 480
 ccattggtcgc gagttagctt gcactggccg tgcgtttaca acgtcgtgac tgggaaaacc 540
 ctggcggttac ccaacttaat cgccttgca ccatccccc tttcgccagc tggcgtaata 600
 gcgaagagge ccgcaccgat cgccttccc aacagttgcy cagcctgaat ggccaatggc 660
 gctttgectg gtttccggca ccagaagcgg tgcgggaaag ctggctggag tgcgatcttc 720
 55 ctgagccga tactgtgtc gtcccctcaa actggcagat gcacggttac gatgcgcca 780
 tctacaccaa cgtaacctat ccattacgg tcaatccgcc gtttgctccc acggagaaac 840
 cgacgggttg ttactcgtc acatttaatg ttgatgaaag ctggctacag gaaggocaga 900
 cgcaattat ttttgatggc gtttaactcg cgtttcatct gtggtgcaac gggcgctggg 960
 tcggttacgg ccaggacagt cgtttgcgt ctgaatttga cctgagcgca tttttacggc 1020
 ccggagaaaa ccgcctcgcg gtgaggtgag tgcgttggag tgacggcagt tatctggaag 1080
 60 atcaggatat tggcggtat agcggcattt tccgtgacgt ctcgctttaa tgatgatttc agccgcgctg 1200
 ctacacaaat cagcgatttc catgttgcca ctgcctttaa tgatgatttc agccgcgctg 1200
 tactggaggc tgaagttcag atgtgcggcg agttgcgtga ctacctacgg gtaacagttt 1260
 ctttatggca gggtgaaacg caggctcgcca gcggcaccgc gcctttcggc ggtgaaatta 1320
 65 tcgatgagcg tgggtggtat gccgatcgcg tcacactacg tctgaacgct gaaaacccga 1380
 aactgtggag ccgcgaaatc ccgaatctct atcgtgcggg ggttgaactg cacacgcgcg 1440
 acggcagct gattgaagca gaagcctgcg atcgtcggtt ccgcgaggtg cggattgaaa 1500
 atggctctgt gctgtgaac ggcaagccgt tgctgattcg aggcgttaac cgtcacgagc 1560

	atcatcctct	gcatgggtcag	gtcatggatg	agcagacgat	ggtgcaggat	atcctgctga	1620
	tgaagcagaa	caactttaac	gcogtgcgt	gttcgcatta	tccgaaccat	ccgctgtggt	1680
	acacgctgtg	cgaccgttac	ggcctgtatg	ttggtggatga	agccaatatt	gaaacccacg	1740
5	gcatgggtgcc	aatgaatcgt	ctgaaccgatg	atccgcgctg	gctaccggcg	atgagcgaac	1800
	gcgtaacgcg	aatgggtgcag	cgcgatogta	atcaccogag	tgtgatcatc	tggtcgctgg	1860
	ggaatgaatc	aggccacggc	gctaatacag	acgcgctgta	tcgctggatc	aaatctgtcg	1920
	atccttcccg	cccggtgcag	tatgaaggcg	gcggagccga	caccacggcc	accgatatta	1980
	tttgcccgat	gtacgcgcgc	gtggatgaag	accagccctt	cccggtgtg	ccgaaatggt	2040
10	ccatcaaaaa	atggctttcg	ctacctggag	agacgcgccc	gctgatcctt	tcggaatacg	2100
	cccacgcgat	gggtaacagt	cttggcggtt	tcgctaaata	ctggcaggcg	tttctgtcagt	2160
	atccccgttt	acagggcggc	ttcgtctggg	actgggtgga	tcagtgcgtg	attaaatatg	2220
	atgaaaacgg	caaccctgtg	tcggcttacg	gcggtgattt	tgccgatacg	ccgaacgatc	2280
	gccagttctg	tatgaacggg	ctgggtcttg	ccgaccgcac	gccgcattcca	gcgctgacgg	2340
15	aagcaaaaaa	ccagcagcag	tttttccagt	tcogtttata	cgggcaaaacc	atcgaagtga	2400
	ccagcgaata	cctgttccgt	catagcgata	acgagctcct	gcactggatg	gtggcgctgg	2460
	atggtaagcc	gctggcaagc	ggtgaagtgc	ctctggatgt	cgctccacaa	ggtaaacagt	2520
	tgattgaact	gcctgaacta	ccgcagccgg	agagcgccgg	gcaactctgg	ctcacagtac	2580
	gcgtagtgca	accgaacggc	acgcgatggt	cagaagccgg	gcacatcagc	gcctggcagc	2640
20	agtggcgctc	ggcggaanaa	ctcagtgatg	cgctccccgc	cgcgctccac	gccatcccg	2700
	atctgaccac	cagcgaaatg	gattttttgca	tcagatggg	taataagcgt	tggcaattta	2760
	accgccagtc	aggctttctt	tcacagatgt	ggattggcga	taaaaaacaa	ctgctgacgc	2820
	cgctgcgcga	tcagttcacc	cgtgcaccgc	tggataacga	cattggcgta	agtgaagcga	2880
	cccgcattha	ccctaaccgc	tgggtcgaac	gctggaagcc	ggcgggccat	taccaggccg	2940
25	aagcagcgtt	gttgcagtgc	acggcagata	caottgotga	tgccgtgctg	attacgaccg	3000
	ctcacgcgtg	gcagcatcag	gggaaaacct	tattttacag	ccggaaaacc	taccggattg	3060
	atggtagtcg	tcacactggc	attacogttg	tgtttgaagt	ggcgagcgat	acaccgcata	3120
	cgccgcggat	tggcctgaac	tgccagctgg	cgaggtagc	agagcgggta	aactggctcg	3180
	gattagggcc	gcaagaaaac	tatcccgacc	gccttactgc	cgctgtttt	gaccgctggg	3240
	atctgccatt	gtcagacatg	tataccccgt	acgtcttccc	gagcgaaaac	ggtctgcgct	3300
30	gcgggacgcg	cgaattgaat	tatggcccac	accagtggcg	cgccgacttc	cagttcaaca	3360
	tcagccgcta	cagtcacacg	caactgatgg	aaaccagcca	tcgccatctg	ctgcacgcgg	3420
	aagaaggcac	atggctgaat	atcgacggtt	tcocataggg	gattgggtgc	gacgactcct	3480
	ggagcccgtc	agtatcgccg	gaattccagc	tgagcgccgg	tcgctaccat	taccagttgg	3540
35	tctgggtgtca	aaaataataa	taaccgggca	ggggggatct	ttgtgaagga	accttacttc	3600
	tgtgggtgtga	cataatttga	caaactacct	acagagattt	aaagctctaa	ggtaaatata	3660
	aaatttttaa	gtgtataatg	tgttaaacta	ctgattctaa	ttgtttgtgt	attttagatt	3720
	ccaacctatg	gaactgatga	atgggacgag	tgggtgaatg	ccagatccag	acatgataag	3780
	atacattgat	gagtttggac	aaaccacaac	tgaattgcag	tgaaaaaaat	gottttattg	3840
40	tgaaatttgt	gatgctattg	ctttattttg	aaccattata	agctgcaata	aacaagttta	3900
	caacaacaat	tgcattcatt	ttatgtttca	ggttcagggg	gaggtgtggg	aggtttttta	3960
	aagcaagtaa	aacctctaca	aatgtggtat	ggctgattat	gatctgcggc	cgcaggccct	4020
	cgtgatacgc	ctatttttat	aggttaatgt	cctgataata	atggtttctt	agacgtcagg	4080
	tggcactttt	cggggaaatg	tgccggaac	ccotatttgt	ttatttttct	aaatacatte	4140
45	aaatatgtat	ccgctcatga	gacaataacc	ctgataaatg	cttcaataat	attgaaaaag	4200
	gaagagtatg	agtattcaac	atttcctgtg	cgcccttatt	ccottttttg	cggcattttg	4260
	ccttcctgtt	tttgctcacc	cagaaacgct	ggtgaaagta	aaagatgctg	aagatcagtt	4320
	gggtgcacga	gtgggttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	4380
	tcgccccgaa	gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	4440
50	attatcccgat	attgacggcg	ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	4500
	tgacttggtt	gagtactcac	cagtcacaga	aaagcatott	acggatggca	tgacagtaag	4560
	agaattatgc	agtgtgcca	taaccatgag	tgataacact	goggccaact	tactttctgac	4620
	aacgatcgga	ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	4680
	tcgccttgat	cgttgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	4740
55	cacgatgcct	gtagcaatgg	caacaacgtt	gcgcaacta	ttactaggcg	aactacttac	4800
	tctagcttcc	cggcaacaat	taatagactg	gtggaggcg	gataaagttg	caggaccact	4860
	tctgcgctcg	gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	4920
	tgggtctcgc	ggtatcattg	cagcaactgg	gccagatggt	aagccctccc	gtatcgtagt	4980
	tatctacacg	acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	5040
60	aggtgcctca	ctgattaagc	atttgtaact	gtcagaccaa	gtttactcat	atatacttta	5100
	gattgattta	aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	5160
	tctcatgacc	aaaatccctt	aacgtgagtt	ttcgttccac	tgagcgctcag	accccgtaga	5220
	aaagatcaaa	ggatcttctt	gagatccttt	ttttctgcgc	gtaatctgct	gcttgcaaac	5280
	aaaaaaacca	ccgctaccag	cggtgggttg	tttgccggat	caagagctac	caactctttt	5340
65	tccgaaggta	actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	5400
	gtagttaggc	caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgctaag	5460
	cctgtttacca	gtggctgctg	ccagtgcgga	taagtcgtgt	cttaccgggt	tggactcaag	5520
	acgatagtta	ccggataagg	cgcagcggtc	gggctgaacg	gggggttcgt	gcacacagcc	5580

14

5 cagcttgagg cgaacgacct acacogaact gagataccta cagcgtgagc tatgagaaag 5640
 cgccacgctt ccgaaggga gaaaggcggg caggtatccg gtaagcggca gggcgggaac 5700
 aggagagcgc acgaggggag ttccaggggg aaacgcctgg tatctttata gtccgtgctg 5760
 gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct 5820
 atggaaaaac gccagcaacg cggccttttt acggttccctg gccttttgct ggccttttgc 5880
 tcacatgttc tttcctgctg tatccctga ttctgtggat aaccgtatta ccgcctttga 5940
 gtgagctgat accgctcgcc gcagccgaac gaccgagcgc agcgagtcag tgagcgagga 6000
 agcgggaagag cgccaatac gcaaaccgcc tctcccgccg cgttgccga ttcattaatg 6060
 10 cagctggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg caattaatgt 6120
 gagttagctc actcattagg caccccaggc ttacacttt atgcttccgg ctcgatggtt 6180
 gtgtggaatt gtgagcggat aacaatttca cacaggaaac agctatgacc atgattacgc 6240
 caagctggcg cg 6252

15 <210> 17
 <211> 25
 <212> DNA
 <213> Artificial Sequence

20 <220>
 <223> Description of Artificial Sequence: primer P64-1

<400> 17
 25 tcagcaacca ggctccccag caggc 25

<210> 18
 <211> 27
 <212> DNA
 30 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer 64-4

35 <400> 18
 gacgacagta tcggcctcag gaagatc 27

40 <210> 19
 <211> 840
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence:
 oligonucleotide 80d

<400> 19
 50 ggtaccgagc tcggatcctc tagtaacggc cgccagtgtg ctggaattcg gcttcagcaa 60
 ccaggctccc cagcaggcag aagtatgcaa agcatgcac tcaattagtc agcaaccagg 120
 tgttgaaagt cccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag 180
 tcagcaacca tagtcccgcc cctaactcgg cccatcccgcc ccctaactcc gccagttccc 240
 gccattctc cgcccatggt ctgactaatt ttttttatct atgcagaggc cgaggccgccc 300
 55 tcggcctagg aacagtcgac gacactgcag agacctactt cactaacaac cggtacagtt 360
 cgtggaccag atgggtgagg tggagtaacg gcccggggag cccaaagggt accccagttg 420
 gggcactact cccgaaaacc gcttctggat ccataacttc gtatagcata cattatacga 480
 agttataccg ggccaccatg gtgcgcagta gcttggcact ggggttgctt ttgcgnygtc 540
 gtgactggga aaaccctggc gttacccaac ttaatcgccct tgcagcacat ccccttttcg 600
 ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag ttgcgcagct 660
 60 gaatggcgaa tggcgctttg cctggcttcc ggcaccagaa gcggtgcccg aaagctggct 720
 ggagtgcgat ctctctgagg ccgatactgt cgtcaagccg aattctgcag atatccatca 780
 cactggcggc cgctcgagca tgcattctaga gggccaattc gccctatagt gagtctgatt 840

65 <210> 20
 <211> 1842
 <212> DNA

<213> Bacteriophage phi-C31

<220>

<221> CDS

<222> (1)..(1839)

<400> 20

5	atg	aca	caa	ggg	gtt	gtg	acc	ggg	gtg	gac	acg	tac	gcg	ggg	gct	tac	48
10	Met	Thr	Gln	Gly	Val	Val	Thr	Gly	Val	Asp	Thr	Tyr	Ala	Gly	Ala	Tyr	
	1				5					10					15		
	gac	cgt	cag	tcg	cgc	gag	cgc	gag	aat	tcg	agc	gca	gca	agc	cca	gcg	96
	Asp	Arg	Gln	Ser	Arg	Glu	Arg	Glu	Asn	Ser	Ser	Ala	Ala	Ser	Pro	Ala	
15				20					25					30			
	aca	cag	cgt	agc	gcc	aac	gaa	gac	aag	gcg	gcc	gac	ctt	cag	cgc	gaa	144
	Thr	Gln	Arg	Ser	Ala	Asn	Glu	Asp	Lys	Ala	Ala	Asp	Leu	Gln	Arg	Glu	
			35					40					45				
20	gtc	gag	cgc	gac	ggg	ggc	cgg	ttc	agg	ttc	gtc	ggg	cat	ttc	agc	gaa	192
	Val	Glu	Arg	Asp	Gly	Gly	Arg	Phe	Arg	Phe	Val	Gly	His	Phe	Ser	Glu	
		50					55					60					
	gcg	ccg	ggc	acg	tcg	ggc	ttc	ggg	acg	gcg	gag	cgc	ccg	gag	ttc	gaa	240
25	Ala	Pro	Gly	Thr	Ser	Ala	Phe	Gly	Thr	Ala	Glu	Arg	Pro	Glu	Phe	Glu	
	65					70				75						80	
	cgc	atc	ctg	aac	gaa	tgc	cgc	gcc	ggg	cgg	ctc	aac	atg	atc	att	gtc	288
30	Arg	Ile	Leu	Asn	Glu	Cys	Arg	Ala	Gly	Arg	Leu	Asn	Met	Ile	Ile	Val	
					85					90					95		
	tat	gac	gtg	tcg	cgc	ttc	tcg	cgc	ctg	aag	gtc	atg	gac	gcg	att	ccg	336
	Tyr	Asp	Val	Ser	Arg	Phe	Ser	Arg	Leu	Lys	Val	Met	Asp	Ala	Ile	Pro	
				100					105					110			
35	att	gtc	tcg	gaa	ttg	ctc	gcc	ctg	ggc	gtg	acg	att	gtt	tcc	act	cag	384
	Ile	Val	Ser	Glu	Leu	Leu	Ala	Leu	Gly	Val	Thr	Ile	Val	Ser	Thr	Gln	
			115					120					125				
40	gaa	ggc	gtc	ttc	cgg	cag	gga	aac	gtc	atg	gac	ctg	att	cac	ctg	att	432
	Glu	Gly	Val	Phe	Arg	Gln	Gly	Asn	Val	Met	Asp	Leu	Ile	His	Leu	Ile	
		130					135					140					
	atg	cgg	ctc	gac	gcg	tcg	cac	aaa	gaa	tct	tcg	ctg	aag	tcg	gcg	aag	480
45	Met	Arg	Leu	Asp	Ala	Ser	His	Lys	Glu	Ser	Ser	Leu	Lys	Ser	Ala	Lys	
	145					150					155					160	
	att	ctc	gac	acg	aag	aac	ctt	cag	cgc	gaa	ttg	ggc	ggg	tac	gtc	ggc	528
50	Ile	Leu	Asp	Thr	Lys	Asn	Leu	Gln	Arg	Glu	Leu	Gly	Gly	Tyr	Val	Gly	
					165					170					175		
	ggg	aag	gcg	cct	tac	ggc	ttc	gag	ott	gtt	tcg	gag	acg	aag	gag	atc	576
	Gly	Lys	Ala	Pro	Tyr	Gly	Phe	Glu	Leu	Val	Ser	Glu	Thr	Lys	Glu	Ile	
				180					185					190			
55	acg	cgc	aac	ggc	cga	atg	gtc	aat	gtc	gtc	atc	aac	aag	ctt	gcg	cac	624
	Thr	Arg	Asn	Gly	Arg	Met	Val	Asn	Val	Val	Ile	Asn	Lys	Leu	Ala	His	
			195					200					205				
60	tcg	acc	act	ccc	ctt	acc	gga	ccc	ttc	gag	ttc	gag	ccc	gac	gta	atc	672
	Ser	Thr	Thr	Pro	Leu	Thr	Gly	Pro	Phe	Glu	Phe	Glu	Pro	Asp	Val	Ile	
		210					215					220					
	cgg	tgg	tgg	tgg	cgt	gag	atc	aag	acg	cac	aaa	cac	ctt	ccc	ttc	aag	720
65	Arg	Trp	Trp	Trp	Arg	Glu	Ile	Lys	Thr	His	Lys	His	Leu	Pro	Phe	Lys	
	225					230					235					240	

	ccg	ggc	agt	caa	gcc	gcc	att	cac	ccg	ggc	agc	atc	acg	ggg	ctt	tgt	768
	Pro	Gly	Ser	Gln	Ala	Ala	Ile	His	Pro	Gly	Ser	Ile	Thr	Gly	Leu	Cys	
					245					250					255		
5	aag	cgc	atg	gac	gct	gac	gcc	gtg	ccg	acc	cgg	ggc	gag	acg	att	ggg	816
	Lys	Arg	Met	Asp	Ala	Asp	Ala	Val	Pro	Thr	Arg	Gly	Glu	Thr	Ile	Gly	
				260					265					270			
10	aag	aag	acc	gct	tca	agc	gcc	tgg	gac	ccg	gca	acc	gtt	atg	cga	atc	864
	Lys	Lys	Thr	Ala	Ser	Ser	Ala	Trp	Asp	Pro	Ala	Thr	Val	Met	Arg	Ile	
			275					280					285				
15	ctt	cgg	gac	ccg	cgt	att	ggc	ggc	ttc	gcc	gct	gag	gtg	atc	tac	aag	912
	Leu	Arg	Asp	Pro	Arg	Ile	Ala	Gly	Phe	Ala	Ala	Glu	Val	Ile	Tyr	Lys	
		290					295					300					
20	aag	aag	ccg	gac	ggc	acg	ccg	acc	acg	aag	att	gag	ggt	tac	cgc	att	960
	Lys	Lys	Pro	Asp	Gly	Thr	Pro	Thr	Thr	Lys	Ile	Glu	Gly	Tyr	Arg	Ile	
	305					310					315					320	
25	cag	cgc	gac	ccg	atc	acg	ctc	cgg	ccg	gtc	gag	ctt	gat	tgc	gga	ccg	1008
	Gln	Arg	Asp	Pro	Ile	Thr	Leu	Arg	Pro	Val	Glu	Leu	Asp	Cys	Gly	Pro	
					325					330					335		
30	atc	atc	gag	ccc	gct	gag	tgg	tat	gag	ctt	cag	gcg	tgg	ttg	gac	ggc	1056
	Ile	Ile	Glu	Pro	Ala	Glu	Trp	Tyr	Glu	Leu	Gln	Ala	Trp	Leu	Asp	Gly	
				340					345					350			
35	agg	ggg	cgc	ggc	aag	ggg	ctt	tcc	cgg	ggg	caa	gcc	att	ctg	tcc	gcc	1104
	Arg	Gly	Arg	Gly	Lys	Gly	Leu	Ser	Arg	Gly	Gln	Ala	Ile	Leu	Ser	Ala	
			355					360					365				
40	atg	gac	aag	ctg	tac	tgc	gag	tgt	ggc	gcc	gtc	atg	act	tgc	aag	cgc	1152
	Met	Asp	Lys	Leu	Tyr	Cys	Glu	Cys	Gly	Ala	Val	Met	Thr	Ser	Lys	Arg	
		370					375					380					
45	ggg	gaa	gaa	tgc	atc	aag	gac	tct	tac	cgc	tgc	cgt	cgc	cgg	aag	gtg	1200
	Gly	Glu	Glu	Ser	Ile	Lys	Asp	Ser	Tyr	Arg	Cys	Arg	Arg	Arg	Lys	Val	
	385					390					395					400	
50	gtc	gac	ccg	tcc	gca	cct	ggg	cag	cac	gaa	ggc	acg	tgc	aac	gtc	agc	1248
	Val	Asp	Pro	Ser	Ala	Pro	Gly	Gln	His	Glu	Gly	Thr	Cys	Asn	Val	Ser	
					405					410					415		
55	atg	gcg	gca	ctc	gac	aag	ttc	gtt	gcg	gaa	cgc	atc	ttc	aac	aag	atc	1296
	Met	Ala	Ala	Leu	Asp	Lys	Phe	Val	Ala	Glu	Arg	Ile	Phe	Asn	Lys	Ile	
				420				425					430				
60	agg	cac	gcc	gaa	ggc	gac	gaa	gag	acg	ttg	gcg	ctt	ctg	tgg	gaa	gcc	1344
	Arg	His	Ala	Glu	Gly	Asp	Glu	Glu	Thr	Leu	Ala	Leu	Leu	Trp	Glu	Ala	
			435				440					445					
65	gcc	cga	cgc	ttc	ggc	aag	ctc	act	gag	gcg	cct	gag	aag	agc	ggc	gaa	1392
	Ala	Arg	Arg	Phe	Gly	Lys	Leu	Thr	Glu	Ala	Pro	Glu	Lys	Ser	Gly	Glu	
		450					455					460					
70	cgg	gcg	aac	ctt	gtt	gcg	gag	cgc	gcc	gac	gcc	ctg	aac	gcc	ctt	gaa	1440
	Arg	Ala	Asn	Leu	Val	Ala	Glu	Arg	Ala	Asp	Ala	Leu	Asn	Ala	Leu	Glu	
	465				470						475					480	
75	gag	ctg	tac	gaa	gac	cgc	gcg	gca	ggc	gcg	tac	gac	gga	ccc	gtt	ggc	1488
	Glu	Leu	Tyr	Glu	Asp	Arg	Ala	Ala	Gly	Ala	Tyr	Asp	Gly	Pro	Val	Gly	
				485					490						495		
80	agg	aag	cac	ttc	cgg	aag	caa	cag	gca	gcg	ctg	acg	ctc	cgg	cag	caa	1536
	Arg	Lys	His	Phe	Arg	Lys	Gln	Gln	Ala	Ala	Leu	Thr	Leu	Arg	Gln	Gln	
				500					505					510			

	ggg gcg gaa gag cgg ctt gcc gaa ctt gaa gcc gcc gaa gcc ccg aag Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Glu Ala Pro Lys	1584
5	515 520 525	
	ctt ccc ctt gac caa tgg ttc ccc gaa gac gcc gac gct gac ccg acc Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr	1632
	530 535 540	
10	ggc cct aag tcg tgg tgg ggg cgc gcg tca gta gac gac aag cgc gtg Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val	1680
	545 550 555 560	
15	ttc gtc ggg ctc ttc gta gac aag atc gtt gtc acg aag tcg act acg Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr	1728
	565 570 575	
20	ggc agg ggg cag gga acg ccc atc gag aag cgc gct tcg atc acg tgg Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp	1776
	580 585 590	
	gcg aag ccg ccg acc gac gac gac gaa gac gac gcc cag gac ggc acg Ala Lys Pro Pro Thr Asp Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr	1824
	595 600 605	
25	gaa gac gta gcg gcg tag Glu Asp Val Ala Ala	1842
	610	
30	<210> 21 <211> 613 <212> PRT <213> Bacteriophage phi-C31	
35	<400> 21 Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr 1 5 10 15	
40	Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala 20 25 30	
	Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu 35 40 45	
45	Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu 50 55 60	
50	Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu 65 70 75 80	
	Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val 85 90 95	
55	Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro 100 105 110	
	Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln 115 120 125	
60	Glu Gly Val Phe Arg Gln Gly Asn Val Met Asp Leu Ile His Leu Ile 130 135 140	
65	Met Arg Leu Asp Ala Ser His Lys Glu Ser Ser Leu Lys Ser Ala Lys 145 150 155 160	
	Ile Leu Asp Thr Lys Asn Leu Gln Arg Glu Leu Gly Gly Tyr Val Gly	

	18															
	165								170				175			
5	Gly	Lys	Ala	Pro	Tyr	Gly	Phe	Glu	Leu	Val	Ser	Glu	Thr	Lys	Glu	Ile
				180					185					190		
	Thr	Arg	Asn	Gly	Arg	Met	Val	Asn	Val	Val	Ile	Asn	Lys	Leu	Ala	His
			195					200					205			
10	Ser	Thr	Thr	Pro	Leu	Thr	Gly	Pro	Phe	Glu	Phe	Glu	Pro	Asp	Val	Ile
		210					215					220				
	Arg	Trp	Trp	Trp	Arg	Glu	Ile	Lys	Thr	His	Lys	His	Leu	Pro	Phe	Lys
	225					230					235					240
15	Pro	Gly	Ser	Gln	Ala	Ala	Ile	His	Pro	Gly	Ser	Ile	Thr	Gly	Leu	Cys
					245					250					255	
	Lys	Arg	Met	Asp	Ala	Asp	Ala	Val	Pro	Thr	Arg	Gly	Glu	Thr	Ile	Gly
				260					265					270		
20	Lys	Lys	Thr	Ala	Ser	Ser	Ala	Trp	Asp	Pro	Ala	Thr	Val	Met	Arg	Ile
			275					280					285			
25	Leu	Arg	Asp	Pro	Arg	Ile	Ala	Gly	Phe	Ala	Ala	Glu	Val	Ile	Tyr	Lys
	290						295					300				
	Lys	Lys	Pro	Asp	Gly	Thr	Pro	Thr	Thr	Lys	Ile	Glu	Gly	Tyr	Arg	Ile
	305					310					315					320
30	Gln	Arg	Asp	Pro	Ile	Thr	Leu	Arg	Pro	Val	Glu	Leu	Asp	Cys	Gly	Pro
					325					330					335	
	Ile	Ile	Glu	Pro	Ala	Glu	Trp	Tyr	Glu	Leu	Gln	Ala	Trp	Leu	Asp	Gly
				340					345					350		
35	Arg	Gly	Arg	Gly	Lys	Gly	Leu	Ser	Arg	Gly	Gln	Ala	Ile	Leu	Ser	Ala
			355					360					365			
40	Met	Asp	Lys	Leu	Tyr	Cys	Glu	Cys	Gly	Ala	Val	Met	Thr	Ser	Lys	Arg
	370						375					380				
	Gly	Glu	Glu	Ser	Ile	Lys	Asp	Ser	Tyr	Arg	Cys	Arg	Arg	Arg	Lys	Val
	385					390					395					400
45	Val	Asp	Pro	Ser	Ala	Pro	Gly	Gln	His	Glu	Gly	Thr	Cys	Asn	Val	Ser
					405					410					415	
	Met	Ala	Ala	Leu	Asp	Lys	Phe	Val	Ala	Glu	Arg	Ile	Phe	Asn	Lys	Ile
				420					425					430		
50	Arg	His	Ala	Glu	Gly	Asp	Glu	Glu	Thr	Leu	Ala	Leu	Leu	Trp	Glu	Ala
			435					440					445			
55	Ala	Arg	Arg	Phe	Gly	Lys	Leu	Thr	Glu	Ala	Pro	Glu	Lys	Ser	Gly	Glu
	450						455					460				
	Arg	Ala	Asn	Leu	Val	Ala	Glu	Arg	Ala	Asp	Ala	Leu	Asn	Ala	Leu	Glu
	465					470				475						480
60	Glu	Leu	Tyr	Glu	Asp	Arg	Ala	Ala	Gly	Ala	Tyr	Asp	Gly	Pro	Val	Gly
					485					490					495	
	Arg	Lys	His	Phe	Arg	Lys	Gln	Gln	Ala	Ala	Leu	Thr	Leu	Arg	Gln	Gln
				500					505					510		
65	Gly	Ala	Glu	Glu	Arg	Leu	Ala	Glu	Leu	Glu	Ala	Ala	Glu	Ala	Pro	Lys
			515					520					525			

Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr
 530 535 540
 5 Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val
 545 550 555 560
 Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr
 565 570 575
 10 Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp
 580 585 590
 15 Ala Lys Pro Pro Thr Asp Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr
 595 600 605
 Glu Asp Val Ala Ala
 610
 20
 <210> 22
 <211> 1863
 <212> DNA
 25 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein C31-Int(CNLS)
 30
 <220>
 <221> CDS
 <222> (1)..(1860)
 35 <400> 22
 atg aca caa ggg gtt gtg acc ggg gtg gac acg tac gcg ggt gct tac 48
 Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr
 1 5 10 15
 40 gac cgt cag tcg cgc gag cgc gag aat tcg agc gca gca agc cca gcg 96
 Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala
 20 25 30
 45 aca cag cgt agc gcc aac gaa gac aag gcg gcc gac ctt cag cgc gaa 144
 Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu
 35 40 45
 50 gtc gag cgc gac ggg ggc cgg ttc agg ttc gtc ggg cat ttc agc gaa 192
 Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu
 50 55 60
 gcg ccg ggc acg tcg gcg ttc ggg acg gcg gag cgc ccg gag ttc gaa 240
 Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu
 65 70 75 80
 55 cgc atc ctg aac gaa tgc cgc gcc ggg cgg ctc aac atg atc att gtc 288
 Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val
 85 90 95
 60 tat gac gtg tcg cgc ttc tcg cgc ctg aag gtc atg gac gcg att ccg 336
 Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro
 100 105 110
 65 att gtc tcg gaa ttg ctc gcc ctg ggc gtg acg att gtt tcc act cag 384
 Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln
 115 120 125

20

	gaa	ggc	gtc	ttc	cgg	cag	gga	aac	gtc	atg	gac	ctg	att	cac	ctg	att	432
	Glu	Gly	Val	Phe	Arg	Gln	Gly	Asn	Val	Met	Asp	Leu	Ile	His	Leu	Ile	
	130						135					140					
5	atg	cgg	ctc	gac	gcg	tgc	cac	aaa	gaa	tct	tgc	ctg	aag	tgc	gcg	aag	480
	Met	Arg	Leu	Asp	Ala	Ser	His	Lys	Glu	Ser	Ser	Leu	Lys	Ser	Ala	Lys	
	145					150					155					160	
10	att	ctc	gac	acg	aag	aac	ctt	cag	cgc	gaa	ttg	ggc	ggg	tac	gtc	ggc	528
	Ile	Leu	Asp	Thr	Lys	Asn	Leu	Gln	Arg	Glu	Leu	Gly	Gly	Tyr	Val	Gly	
					165					170					175		
15	ggg	aag	gcg	cct	tac	ggc	ttc	gag	ctt	gtt	tgc	gag	acg	aag	gag	atc	576
	Gly	Lys	Ala	Pro	Tyr	Gly	Phe	Glu	Leu	Val	Ser	Glu	Thr	Lys	Glu	Ile	
				180					185					190			
20	acg	cgc	aac	ggc	cga	atg	gtc	aat	gtc	gtc	atc	aac	aag	ctt	gcg	cac	624
	Thr	Arg	Asn	Gly	Arg	Met	Val	Asn	Val	Val	Ile	Asn	Lys	Leu	Ala	His	
				195				200					205				
25	tgc	acc	act	ccc	ctt	acc	gga	ccc	ttc	gag	ttc	gag	ccc	gac	gta	atc	672
	Ser	Thr	Thr	Pro	Leu	Thr	Gly	Pro	Phe	Glu	Phe	Glu	Pro	Asp	Val	Ile	
		210					215					220					
30	cgg	tgg	tgg	tgg	cgt	gag	atc	aag	acg	cac	aaa	cac	ctt	ccc	ttc	aag	720
	Arg	Trp	Trp	Trp	Arg	Glu	Ile	Lys	Thr	His	Lys	His	Leu	Pro	Phe	Lys	
	225					230					235					240	
35	ccg	ggc	agt	caa	goc	goc	att	cac	ccg	ggc	agc	atc	acg	ggg	ctt	tgt	768
	Pro	Gly	Ser	Gln	Ala	Ala	Ile	His	Pro	Gly	Ser	Ile	Thr	Gly	Leu	Cys	
					245					250					255		
40	aag	cgc	atg	gac	gct	gac	gcc	gtg	ccg	acc	cgg	ggc	gag	acg	att	ggg	816
	Lys	Arg	Met	Asp	Ala	Asp	Ala	Val	Pro	Thr	Arg	Gly	Glu	Thr	Ile	Gly	
				260					265					270			
45	aag	aag	acc	gct	tca	agc	gcc	tgg	gac	ccg	gca	acc	gtt	atg	cga	atc	864
	Lys	Lys	Thr	Ala	Ser	Ser	Ala	Trp	Asp	Pro	Ala	Thr	Val	Met	Arg	Ile	
			275					280					285				
50	ctt	cgg	gac	ccg	cgt	att	gcg	ggc	ttc	gcc	gct	gag	gtg	atc	tac	aag	912
	Leu	Arg	Asp	Pro	Arg	Ile	Ala	Gly	Phe	Ala	Ala	Glu	Val	Ile	Tyr	Lys	
		290				295						300					
55	aag	aag	ccg	gac	ggc	acg	ccg	acc	acg	aag	att	gag	ggt	tac	cgc	att	960
	Lys	Lys	Pro	Asp	Gly	Thr	Pro	Thr	Thr	Lys	Ile	Glu	Gly	Tyr	Arg	Ile	
	305					310					315					320	
60	cag	cgc	gac	ccg	atc	acg	ctc	cgg	ccg	gtc	gag	ctt	gat	tgc	gga	ccg	1008
	Gln	Arg	Asp	Pro	Ile	Thr	Leu	Arg	Pro	Val	Glu	Leu	Asp	Cys	Gly	Pro	
					325					330					335		
65	atc	atc	gag	ccc	gct	gag	tgg	tat	gag	ctt	cag	gcg	tgg	ttg	gac	ggc	1056
	Ile	Ile	Glu	Pro	Ala	Glu	Trp	Tyr	Glu	Leu	Gln	Ala	Trp	Leu	Asp	Gly	
				340					345					350			
70	agg	ggg	cgc	ggc	aag	ggg	ctt	tcc	cgg	ggg	caa	gcc	att	ctg	tcc	gcc	1104
	Arg	Gly	Arg	Gly	Lys	Gly	Leu	Ser	Arg	Gly	Gln	Ala	Ile	Leu	Ser	Ala	
				355				360					365				
75	atg	gac	aag	ctg	tac	tgc	gag	tgt	ggc	gcc	gtc	atg	act	tgc	aag	cgc	1152
	Met	Asp	Lys	Leu	Tyr	Cys	Glu	Cys	Gly	Ala	Val	Met	Thr	Ser	Lys	Arg	
		370					375					380					
80	ggg	gaa	gaa	tgc	atc	aag	gac	tct	tac	cgc	tgc	cgt	cgc	cgg	aag	gtg	1200
	Gly	Glu	Glu	Ser	Ile	Lys	Asp	Ser	Tyr	Arg	Cys	Arg	Arg	Arg	Lys	Val	
	385					390					395					400	

21

	gtc gac ccg tcc gca cct ggg cag cac gaa ggc acg tgc aac gtc agc	1248								
	Val Asp Pro Ser Ala Pro Gly Gln His Glu Gly Thr Cys Asn Val Ser									
5	405 410 415									
	atg gcg gca ctc gac aag ttc gtt gcg gaa cgc atc ttc aac aag atc	1296								
	Met Ala Ala Leu Asp Lys Phe Val Ala Glu Arg Ile Phe Asn Lys Ile									
	420 425 430									
10	agg cac gcc gaa ggc gac gaa gag acg ttg gcg ctt ctg tgg gaa gcc	1344								
	Arg His Ala Glu Gly Asp Glu Glu Thr Leu Ala Leu Leu Trp Glu Ala									
	435 440 445									
15	gcc cga cgc ttc ggc aag ctc act gag gcg cct gag aag agc ggc gaa	1392								
	Ala Arg Arg Phe Gly Lys Leu Thr Glu Ala Pro Glu Lys Ser Gly Glu									
	450 455 460									
20	cgg gcg aac ctt gtt gcg gag cgc gcc gac gcc ctg aac gcc ctt gaa	1440								
	Arg Ala Asn Leu Val Ala Glu Arg Ala Asp Ala Leu Asn Ala Leu Glu									
	465 470 475 480									
25	gag ctg tac gaa gac cgc gcg gca ggc gcg tac gac gga ccc gtt ggc	1488								
	Glu Leu Tyr Glu Asp Arg Ala Ala Gly Ala Tyr Asp Gly Pro Val Gly									
	485 490 495									
	agg aag cac ttc cgg aag caa cag gca gcg ctg acg ctc cgg cag caa	1536								
	Arg Lys His Phe Arg Lys Gln Gln Ala Ala Leu Thr Leu Arg Gln Gln									
	500 505 510									
30	ggg gcg gaa gag cgg ctt gcc gaa ctt gaa gcc gcc gaa gcc ccg aag	1584								
	Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Glu Ala Pro Lys									
	515 520 525									
35	ctt ccc ctt gac caa tgg ttc ccc gaa gac gcc gac gct gac ccg acc	1632								
	Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr									
	530 535 540									
40	ggc cct aag tcg tgg tgg ggg cgc gcg tca gta gac gac aag cgc gtg	1680								
	Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val									
	545 550 555 560									
45	ttc gtc ggg ctc ttc gta gac aag atc gtt gtc acg aag tcg act acg	1728								
	Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr									
	565 570 575									
	ggc agg ggg cag gga acg ccc atc gag aag cgc gct tcg atc acg tgg	1776								
	Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp									
	580 585 590									
50	gcg aag ccg ccg acc gac gac gac gaa gac gac gcc cag gac ggc acg	1824								
	Ala Lys Pro Pro Thr Asp Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr									
	595 600 605									
55	gaa gac gta gcg gcg cct aag aag aag agg aag gtt tag	1863								
	Glu Asp Val Ala Ala Pro Lys Lys Lys Arg Lys Val									
	610 615 620									
60	<210> 23									
	<211> 620									
	<212> PRT									
	<213> Artificial Sequence									
	<223> Description of Artificial Sequence: DNA sequence									
	coding for fusion protein C31-Int(CNLS)									
65	<400> 23									
	Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr									

	1				5				22 10				15			
5	Asp	Arg	Gln	Ser	Arg	Glu	Arg	Glu	Asn	Ser	Ser	Ala	Ala	Ser	Pro	Ala
				20					25					30		
	Thr	Gln	Arg	Ser	Ala	Asn	Glu	Asp	Lys	Ala	Ala	Asp	Leu	Gln	Arg	Glu
			35					40					45			
10	Val	Glu	Arg	Asp	Gly	Gly	Arg	Phe	Arg	Phe	Val	Gly	His	Phe	Ser	Glu
		50					55					60				
	Ala	Pro	Gly	Thr	Ser	Ala	Phe	Gly	Thr	Ala	Glu	Arg	Pro	Glu	Phe	Glu
	65					70					75					80
15	Arg	Ile	Leu	Asn	Glu	Cys	Arg	Ala	Gly	Arg	Leu	Asn	Met	Ile	Ile	Val
				85						90					95	
	Tyr	Asp	Val	Ser	Arg	Phe	Ser	Arg	Leu	Lys	Val	Met	Asp	Ala	Ile	Pro
				100					105					110		
20	Ile	Val	Ser	Glu	Leu	Leu	Ala	Leu	Gly	Val	Thr	Ile	Val	Ser	Thr	Gln
			115					120					125			
	Glu	Gly	Val	Phe	Arg	Gln	Gly	Asn	Val	Met	Asp	Leu	Ile	His	Leu	Ile
25		130					135					140				
	Met	Arg	Leu	Asp	Ala	Ser	His	Lys	Glu	Ser	Ser	Leu	Lys	Ser	Ala	Lys
	145					150					155					160
30	Ile	Leu	Asp	Thr	Lys	Asn	Leu	Gln	Arg	Glu	Leu	Gly	Gly	Tyr	Val	Gly
					165					170					175	
	Gly	Lys	Ala	Pro	Tyr	Gly	Phe	Glu	Leu	Val	Ser	Glu	Thr	Lys	Glu	Ile
				180					185					190		
35	Thr	Arg	Asn	Gly	Arg	Met	Val	Asn	Val	Val	Ile	Asn	Lys	Leu	Ala	His
			195					200					205			
	Ser	Thr	Thr	Pro	Leu	Thr	Gly	Pro	Phe	Glu	Phe	Glu	Pro	Asp	Val	Ile
40		210					215					220				
	Arg	Trp	Trp	Trp	Arg	Glu	Ile	Lys	Thr	His	Lys	His	Leu	Pro	Phe	Lys
	225					230					235					240
45	Pro	Gly	Ser	Gln	Ala	Ala	Ile	His	Pro	Gly	Ser	Ile	Thr	Gly	Leu	Cys
				245						250					255	
	Lys	Arg	Met	Asp	Ala	Asp	Ala	Val	Pro	Thr	Arg	Gly	Glu	Thr	Ile	Gly
				260					265					270		
50	Lys	Lys	Thr	Ala	Ser	Ser	Ala	Trp	Asp	Pro	Ala	Thr	Val	Met	Arg	Ile
			275					280					285			
	Leu	Arg	Asp	Pro	Arg	Ile	Ala	Gly	Phe	Ala	Ala	Glu	Val	Ile	Tyr	Lys
55		290					295					300				
	Lys	Lys	Pro	Asp	Gly	Thr	Pro	Thr	Thr	Lys	Ile	Glu	Gly	Tyr	Arg	Ile
	305					310					315					320
60	Gln	Arg	Asp	Pro	Ile	Thr	Leu	Arg	Pro	Val	Glu	Leu	Asp	Cys	Gly	Pro
					325					330					335	
	Ile	Ile	Glu	Pro	Ala	Glu	Trp	Tyr	Glu	Leu	Gln	Ala	Trp	Leu	Asp	Gly
				340					345					350		
65	Arg	Gly	Arg	Gly	Lys	Gly	Leu	Ser	Arg	Gly	Gln	Ala	Ile	Leu	Ser	Ala
			355					360					365			

Met Asp Lys Leu Tyr Cys Glu Cys Gly Ala Val Met Thr Ser Lys Arg
 370 375 380
 5 Gly Glu Glu Ser Ile Lys Asp Ser Tyr Arg Cys Arg Arg Arg Lys Val
 385 390 395 400
 Val Asp Pro Ser Ala Pro Gly Gln His Glu Gly Thr Cys Asn Val Ser
 405 410 415
 10 Met Ala Ala Leu Asp Lys Phe Val Ala Glu Arg Ile Phe Asn Lys Ile
 420 425 430
 15 Arg His Ala Glu Gly Asp Glu Glu Thr Leu Ala Leu Leu Trp Glu Ala
 435 440 445
 Ala Arg Arg Phe Gly Lys Leu Thr Glu Ala Pro Glu Lys Ser Gly Glu
 450 455 460
 20 Arg Ala Asn Leu Val Ala Glu Arg Ala Asp Ala Leu Asn Ala Leu Glu
 465 470 475 480
 Glu Leu Tyr Glu Asp Arg Ala Ala Gly Ala Tyr Asp Gly Pro Val Gly
 485 490 495
 25 Arg Lys His Phe Arg Lys Gln Gln Ala Ala Leu Thr Leu Arg Gln Gln
 500 505 510
 30 Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Glu Ala Pro Lys
 515 520 525
 Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr
 530 535 540
 35 Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val
 545 550 555 560
 Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr
 565 570 575
 40 Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp
 580 585 590
 45 Ala Lys Pro Pro Thr Asp Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr
 595 600 605
 Glu Asp Val Ala Ala Pro Lys Lys Lys Arg Lys Val
 610 615 620
 50
 <210> 24
 <211> 43
 <212> PRT
 55 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: NLS
 60 <400> 24
 Met Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Cys Arg Leu Lys
 1 5 10 15
 65 Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu Lys
 20 25 30
 Lys Lys Lys Lys Arg Arg Arg Lys Thr Lys Arg

35

40

5 <210> 25
 <211> 10
 <212> PRT
 <213> Artificial Sequence

 10 <220>
 <223> Description of Artificial Sequence: NLS

 <400> 25
 Ile Lys Tyr Phe Lys Lys Phe Pro Lys Asp
 1 5 10
 15

 <210> 26
 <211> 14
 <212> PRT
 20 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: NLS

 25 <400> 26
 Met Thr Gly Ser Lys Thr Arg Lys His Arg Gly Ser Gly Ala
 1 5 10

 30 <210> 27
 <211> 14
 <212> PRT
 <213> Artificial Sequence

 35 <220>
 <223> Description of Artificial Sequence: NLS

 <400> 27
 40 Met Thr Gly Ser Lys His Arg Lys His Pro Gly Ser Gly Ala
 1 5 10

 <210> 28
 <211> 7
 45 <212> PRT
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: NLS
 50
 <400> 28
 Gly Lys Lys Arg Ser Lys Ala
 1 5

 55 <210> 29
 <211> 14
 <212> PRT
 <213> Artificial Sequence
 60
 <220>
 <223> Description of Artificial Sequence: NLS

 <400> 29
 65 Pro Lys Lys Ala Arg Glu Asp Val Ser Arg Lys Arg Pro Arg
 1 5 10

<210> 30
 <211> 11
 <212> PRT
 5 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: NLS
 10 <400> 30
 Ala Pro Lys Arg Lys Ser Gly Val Ser Lys Cys
 1 5 10

 15 <210> 31
 <211> 12
 <212> PRT
 <213> Artificial Sequence

 20 <220>
 <223> Description of Artificial Sequence: NLS

 <400> 31
 25 Glu Glu Asp Gly Pro Gln Lys Lys Lys Arg Arg Leu
 1 5 10

 <210> 32
 30 <211> 8
 <212> PRT
 <213> Artificial Sequence

 <220>
 35 <223> Description of Artificial Sequence: NLS

 <400> 32
 Ala Pro Thr Lys Arg Lys Gly Ser
 1 5

 40 <210> 33
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 45 <220>
 <223> Description of Artificial Sequence: NLS

 <400> 33
 50 Pro Asn Lys Lys Lys Arg Lys
 1 5

 55 <210> 34
 <211> 5
 <212> PRT
 <213> Artificial Sequence

 <220>
 60 <223> Description of Artificial Sequence: NLS

 <400> 34
 Lys Arg Pro Arg Pro
 1 5
 65 <210> 35

<211> 11
 <212> PRT
 <213> Artificial Sequence
 5 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 35
 10 Cys Gly Gly Leu Ser Ser Lys Arg Pro Arg Pro
 1 5 10
 <210> 36
 <211> 19
 15 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: NLS
 20 <400> 36
 Pro Pro Lys Lys Arg Met Arg Arg Arg Ile Glu Pro Lys Lys Lys Lys
 1 5 10 15
 25 Lys Arg Pro
 <210> 37
 30 <211> 11
 <212> PRT
 <213> Artificial Sequence
 <220>
 35 <223> Description of Artificial Sequence: NLS
 <400> 37
 Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys
 1 5 10
 40 <210> 38
 <211> 9
 <212> PRT
 45 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: NLS
 50 <400> 38
 Pro Lys Gln Lys Arg Lys Met Ala Arg
 1 5
 55 <210> 39
 <211> 9
 <212> PRT
 <213> Artificial Sequence
 60 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 39
 65 Ser Val Thr Lys Lys Arg Lys Leu Glu
 1 5

5 <210> 40
 <211> 11
 <212> PRT
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: NLS

10 <400> 40
 Cys Gly Gly Ala Ala Lys Arg Val Lys Leu Asp
 1 5 10

15 <210> 41
 <211> 9
 <212> PRT
 <213> Artificial Sequence

 <220>
20 <223> Description of Artificial Sequence: NLS

 <400> 41
 Pro Ala Ala Lys Arg Val Lys Leu Asp
 1 5

25 <210> 42
 <211> 11
 <212> PRT
30 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: NLS

35 <400> 42
 Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
 1 5 10

40 <210> 43
 <211> 8
 <212> PRT
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: NLS

 <400> 43
50 Pro Gln Ser Arg Lys Lys Leu Arg
 1 5

55 <210> 44
 <211> 8
 <212> PRT
 <213> Artificial Sequence

 <220>
60 <223> Description of Artificial Sequence: NLS

 <400> 44
 Pro Leu Leu Lys Lys Ile Lys Gln
 1 5

65 <210> 45
 <211> 7

<212> PRT
 <213> Artificial Sequence
 <220>
 5 <223> Description of Artificial Sequence: NLS
 <400> 45
 Pro Gln Pro Lys Lys Lys Pro
 1 5
 10
 <210> 46
 <211> 9
 <212> PRT
 15 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: NLS
 20 <400> 46
 Ser Lys Arg Val Ala Lys Arg Lys Leu
 1 5
 25 <210> 47
 <211> 9
 <212> PRT
 <213> Artificial Sequence
 30 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 47
 35 Ala Ser Lys Ser Arg Lys Arg Lys Leu
 1 5
 <210> 48
 <211> 16
 40 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: NLS
 45 <400> 48
 Gly Gly Leu Cys Ser Ala Arg Leu His Arg His Ala Leu Leu Ala Thr
 1 5 10 15
 50 <210> 49
 <211> 8
 <212> PRT
 <213> Artificial Sequence
 55 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 49
 60 Arg Lys Thr Lys Lys Lys Ile Lys
 1 5
 <210> 50
 65 <211> 8
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: NLS
 5 <400> 50
 Arg Lys Leu Lys Lys Leu Gly Asn
 1 5
 10 <210> 51
 <211> 8
 <212> PRT
 <213> Artificial Sequence
 15 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 51
 Arg Lys Asp Arg Arg Gly Gly Arg
 20 1 5
 <210> 52
 <211> 18
 25 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: NLS
 30 <400> 52
 Asp Thr Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg Leu Leu Arg Leu
 1 5 10 15
 35 Asp Glu
 <210> 53
 40 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220>
 45 <223> Description of Artificial Sequence: NLS
 <400> 53
 Pro Lys Lys Lys Arg Lys Val
 1 5
 50 <210> 54
 <211> 1410
 <212> DNA
 55 <213> Bacteriophage R4
 <220>
 <221> CDS
 <222> (1)..(1407)
 60 <400> 54
 atg aat cga ggg ggg ccc act gta cgg gcc gac atc tac gtc cga atc 48
 Met Asn Arg Gly Gly Pro Thr Val Arg Ala Asp Ile Tyr Val Arg Ile
 1 5 10 15
 65 agc ctg gac cgc aca ggg gaa gag ctc ggg gtc gag cgc cag gag gag 96
 Ser Leu Asp Arg Thr Gly Glu Glu Leu Gly Val Glu Arg Gln Glu Glu

	20					25					30						
5	tcg	tgt	cgc	gag	ctc	tgc	aag	agc	ctc	ggc	atg	gag	gtg	ggg	cag	gtg	144
	Ser	Cys	Arg	Glu	Leu	Cys	Lys	Ser	Leu	Gly	Met	Glu	Val	Gly	Gln	Val	
			35					40					45				
10	tgg	gtc	gac	aac	gac	ctg	agc	gcc	acc	aag	aag	aac	gtc	gtc	cgc	cct	192
	Trp	Val	Asp	Asn	Asp	Leu	Ser	Ala	Thr	Lys	Lys	Asn	Val	Val	Arg	Pro	
		50					55					60					
15	gac	ttc	gag	gcg	atg	atc	gcg	agc	aac	ccg	cag	gcg	atc	gtc	tgc	tgg	240
	Asp	Phe	Glu	Ala	Met	Ile	Ala	Ser	Asn	Pro	Gln	Ala	Ile	Val	Cys	Trp	
	65					70					75					80	
20	cac	acc	gac	cgg	ctc	atc	cgc	gtc	acg	cgg	gac	ctg	gag	cgg	gtg	atc	288
	His	Thr	Asp	Arg	Leu	Ile	Arg	Val	Thr	Arg	Asp	Leu	Glu	Arg	Val	Ile	
					85					90					95		
25	gac	ctc	gga	gtc	aac	gtc	cac	gcc	gtg	atg	gcc	gga	cac	ctg	gac	ctg	336
	Asp	Leu	Gly	Val	Asn	Val	His	Ala	Val	Met	Ala	Gly	His	Leu	Asp	Leu	
				100					105					110			
30	tcc	acc	ccg	gcc	ggc	cga	gcc	gtc	gcc	cgc	acg	gtg	acg	gcc	tgg	gcc	384
	Ser	Thr	Pro	Ala	Gly	Arg	Ala	Val	Ala	Arg	Thr	Val	Thr	Ala	Trp	Ala	
			115					120					125				
35	acg	tac	gag	ggc	gag	cag	aag	gct	gag	cgc	cag	aag	ctc	gcc	aac	atc	432
	Thr	Tyr	Glu	Gly	Glu	Gln	Lys	Ala	Glu	Arg	Gln	Lys	Leu	Ala	Asn	Ile	
		130					135					140					
40	cag	aac	gcc	cgc	gcc	ggc	aag	ccg	tac	acc	ccc	ggc	atc	cgc	ccc	ttc	480
	Gln	Asn	Ala	Arg	Ala	Gly	Lys	Pro	Tyr	Thr	Pro	Gly	Ile	Arg	Pro	Phe	
	145					150					155					160	
45	ggg	tac	ggc	gac	gac	cac	atg	acc	atc	gtg	acg	gcc	gag	gcg	gac	gcc	528
	Gly	Tyr	Gly	Asp	Asp	His	Met	Thr	Ile	Val	Thr	Ala	Glu	Ala	Asp	Ala	
					165					170					175		
50	atc	cgc	gac	ggc	gcg	aag	atg	atc	ctc	gac	ggc	tgg	tcc	ctg	tgc	gcc	576
	Ile	Arg	Asp	Gly	Ala	Lys	Met	Ile	Leu	Asp	Gly	Trp	Ser	Leu	Ser	Ala	
				180					185					190			
55	gtg	gct	cgc	tac	tgg	gag	gag	ctc	aag	ctc	cag	tgc	ccc	cgg	agt	atg	624
	Val	Ala	Arg	Tyr	Trp	Glu	Glu	Leu	Lys	Leu	Gln	Ser	Pro	Arg	Ser	Met	
			195					200					205				
60	gcc	gca	ggc	ggc	aag	ggc	tgg	tct	ctg	cgg	ggc	gta	aag	aag	gtg	ctg	672
	Ala	Ala	Gly	Gly	Lys	Gly	Trp	Ser	Leu	Arg	Gly	Val	Lys	Lys	Val	Leu	
		210					215					220					
65	acc	tcc	ccg	cgc	tac	gtc	ggg	cgg	tcc	agc	tac	ctc	ggg	gag	gtc	gtg	720
	Thr	Ser	Pro	Arg	Tyr	Val	Gly	Arg	Ser	Ser	Tyr	Leu	Gly	Glu	Val	Val	
	225					230					235					240	
70	ggc	gat	gct	cag	tgg	ccg	ccc	atc	ctc	gac	ccg	gac	gtc	tac	tac	ggg	768
	Gly	Asp	Ala	Gln	Trp	Pro	Pro	Ile	Leu	Asp	Pro	Asp	Val	Tyr	Tyr	Gly	
					245					250					255		
75	gtc	gtg	gcc	atc	ctg	aac	aac	ccc	gac	cgc	ttc	agc	ggg	ggc	cct	cgg	816
	Val	Val	Ala	Ile	Leu	Asn	Asn	Pro	Asp	Arg	Phe	Ser	Gly	Gly	Pro	Arg	
				260					265					270			
80	acc	ggc	cgc	acc	ccc	ggc	acg	ctg	ctc	gca	ggc	atc	gcc	ttg	tgc	ggg	864
	Thr	Gly	Arg	Thr	Pro	Gly	Thr	Leu	Leu	Ala	Gly	Ile	Ala	Leu	Cys	Gly	
			275					280					285				
85	gag	tgc	ggc	aag	acg	gtc	agt	gga	cgc	ggc	tac	cga	ggg	gtc	ctg	gtc	912

31

	Glu	Cys	Gly	Lys	Thr	Val	Ser	Gly	Arg	Gly	Tyr	Arg	Gly	Val	Leu	Val	
	290						295					300					
5	tac	gga	tgt	aag	gac	acg	cac	act	cgg	acg	cct	cgg	agc	atc	gct	gac	960
	Tyr	Gly	Cys	Lys	Asp	Thr	His	Thr	Arg	Thr	Pro	Arg	Ser	Ile	Ala	Asp	
	305					310					315					320	
10	ggc	cgc	gcg	agc	agc	tcg	acc	ctc	gcc	cgg	ctc	atg	ttc	ccc	gac	ttc	1008
	Gly	Arg	Ala	Ser	Ser	Ser	Thr	Leu	Ala	Arg	Leu	Met	Phe	Pro	Asp	Phe	
					325					330					335		
15	ctg	ccc	ggc	ctc	ctg	gcc	tct	ggg	cag	gcc	gag	gac	ggc	cag	tcg	gca	1056
	Leu	Pro	Gly	Leu	Leu	Ala	Ser	Gly	Gln	Ala	Glu	Asp	Gly	Gln	Ser	Ala	
					340				345					350			
20	gca	tcc	aag	cac	tcg	gag	gcc	cag	acg	ctg	cgc	gag	cgc	ctt	gac	ggg	1104
	Ala	Ser	Lys	His	Ser	Glu	Ala	Gln	Thr	Leu	Arg	Glu	Arg	Leu	Asp	Gly	
			355					360					365				
25	ctg	gct	acg	gcc	tac	gcg	gag	ggg	gcg	atc	agc	ctg	tct	cag	atg	acg	1152
	Leu	Ala	Thr	Ala	Tyr	Ala	Glu	Gly	Ala	Ile	Ser	Leu	Ser	Gln	Met	Thr	
		370					375					380					
30	gcc	ggc	tcg	gaa	gca	ctg	cgg	aag	aag	ctg	gag	gtg	atc	gaa	gcc	gac	1200
	Ala	Gly	Ser	Glu	Ala	Leu	Arg	Lys	Lys	Leu	Glu	Val	Ile	Glu	Ala	Asp	
	385					390				395						400	
35	ctc	gtg	ggc	tcg	gca	ggc	atc	ccg	ccc	ttc	gat	cca	gtg	gcc	gga	gtg	1248
	Leu	Val	Gly	Ser	Ala	Gly	Ile	Pro	Pro	Phe	Asp	Pro	Val	Ala	Gly	Val	
					405					410					415		
40	gct	ggc	ctg	atc	tcc	ggc	tgg	ccc	acc	acg	cct	ctc	ccg	acg	cgt	cga	1296
	Ala	Gly	Leu	Ile	Ser	Gly	Trp	Pro	Thr	Thr	Pro	Leu	Pro	Thr	Arg	Arg	
					420				425					430			
45	gca	tgg	gtg	gac	ttc	tgc	ctg	gtg	gtc	acg	ctg	aac	acc	cag	aag	ggg	1344
	Ala	Trp	Val	Asp	Phe	Cys	Leu	Val	Val	Thr	Leu	Asn	Thr	Gln	Lys	Gly	
		435					440						445				
50	cgc	cat	gcg	tcg	agc	atg	acc	gtg	gac	gac	cac	gtc	acc	atc	gag	tgg	1392
	Arg	His	Ala	Ser	Ser	Met	Thr	Val	Asp	Asp	His	Val	Thr	Ile	Glu	Trp	
		450					455					460					
55	cga	gac	gtg	gcc	gag	tag											1410
	Arg	Asp	Val	Ala	Glu												
	465																
60	<210>	55															
	<211>	469															
	<212>	PRT															
	<213>	Bacteriophage R4															
65	<400>	55															
	Met	Asn	Arg	Gly	Gly	Pro	Thr	Val	Arg	Ala	Asp	Ile	Tyr	Val	Arg	Ile	
	1				5					10					15		
70	Ser	Leu	Asp	Arg	Thr	Gly	Glu	Glu	Leu	Gly	Val	Glu	Arg	Gln	Glu	Glu	
				20					25					30			
75	Ser	Cys	Arg	Glu	Leu	Cys	Lys	Ser	Leu	Gly	Met	Glu	Val	Gly	Gln	Val	
			35					40					45				
80	Trp	Val	Asp	Asn	Asp	Leu	Ser	Ala	Thr	Lys	Lys	Asn	Val	Val	Arg	Pro	
		50					55					60					
85	Asp	Phe	Glu	Ala	Met	Ile	Ala	Ser	Asn	Pro	Gln	Ala	Ile	Val	Cys	Trp	

	32															
	65				70				75				80			
5	His	Thr	Asp	Arg	Leu	Ile	Arg	Val	Thr	Arg	Asp	Leu	Glu	Arg	Val	Ile
					85					90					95	
	Asp	Leu	Gly	Val	Asn	Val	His	Ala	Val	Met	Ala	Gly	His	Leu	Asp	Leu
				100					105					110		
10	Ser	Thr	Pro	Ala	Gly	Arg	Ala	Val	Ala	Arg	Thr	Val	Thr	Ala	Trp	Ala
			115					120					125			
	Thr	Tyr	Glu	Gly	Glu	Gln	Lys	Ala	Glu	Arg	Gln	Lys	Leu	Ala	Asn	Ile
		130					135					140				
15	Gln	Asn	Ala	Arg	Ala	Gly	Lys	Pro	Tyr	Thr	Pro	Gly	Ile	Arg	Pro	Phe
	145					150					155					160
	Gly	Tyr	Gly	Asp	Asp	His	Met	Thr	Ile	Val	Thr	Ala	Glu	Ala	Asp	Ala
				165						170					175	
20	Ile	Arg	Asp	Gly	Ala	Lys	Met	Ile	Leu	Asp	Gly	Trp	Ser	Leu	Ser	Ala
				180					185					190		
25	Val	Ala	Arg	Tyr	Trp	Glu	Glu	Leu	Lys	Leu	Gln	Ser	Pro	Arg	Ser	Met
			195					200					205			
	Ala	Ala	Gly	Gly	Lys	Gly	Trp	Ser	Leu	Arg	Gly	Val	Lys	Lys	Val	Leu
		210					215					220				
30	Thr	Ser	Pro	Arg	Tyr	Val	Gly	Arg	Ser	Ser	Tyr	Leu	Gly	Glu	Val	Val
	225					230					235					240
	Gly	Asp	Ala	Gln	Trp	Pro	Pro	Ile	Leu	Asp	Pro	Asp	Val	Tyr	Tyr	Gly
				245						250					255	
35	Val	Val	Ala	Ile	Leu	Asn	Asn	Pro	Asp	Arg	Phe	Ser	Gly	Gly	Pro	Arg
				260					265					270		
40	Thr	Gly	Arg	Thr	Pro	Gly	Thr	Leu	Leu	Ala	Gly	Ile	Ala	Leu	Cys	Gly
			275					280					285			
	Glu	Cys	Gly	Lys	Thr	Val	Ser	Gly	Arg	Gly	Tyr	Arg	Gly	Val	Leu	Val
		290					295					300				
45	Tyr	Gly	Cys	Lys	Asp	Thr	His	Thr	Arg	Thr	Pro	Arg	Ser	Ile	Ala	Asp
	305					310					315					320
	Gly	Arg	Ala	Ser	Ser	Ser	Thr	Leu	Ala	Arg	Leu	Met	Phe	Pro	Asp	Phe
				325						330					335	
50	Leu	Pro	Gly	Leu	Leu	Ala	Ser	Gly	Gln	Ala	Glu	Asp	Gly	Gln	Ser	Ala
				340					345					350		
55	Ala	Ser	Lys	His	Ser	Glu	Ala	Gln	Thr	Leu	Arg	Glu	Arg	Leu	Asp	Gly
			355					360					365			
	Leu	Ala	Thr	Ala	Tyr	Ala	Glu	Gly	Ala	Ile	Ser	Leu	Ser	Gln	Met	Thr
		370					375					380				
60	Ala	Gly	Ser	Glu	Ala	Leu	Arg	Lys	Lys	Leu	Glu	Val	Ile	Glu	Ala	Asp
	385					390					395					400
	Leu	Val	Gly	Ser	Ala	Gly	Ile	Pro	Pro	Phe	Asp	Pro	Val	Ala	Gly	Val
				405						410					415	
65	Ala	Gly	Leu	Ile	Ser	Gly	Trp	Pro	Thr	Thr	Pro	Leu	Pro	Thr	Arg	Arg
				420					425					430		

	Ala	Trp	Val	Asp	Phe	Cys	Leu	Val	Val	Thr	Leu	Asn	Thr	Gln	Lys	Gly	
			435					440					445				
5	Arg	His	Ala	Ser	Ser	Met	Thr	Val	Asp	Asp	His	Val	Thr	Ile	Glu	Trp	
		450					455					460					
10	Arg	Asp	Val	Ala	Glu												
	465																
	<210>	56															
	<211>	1503															
15	<212>	DNA															
	<213>	CisA recombinase															
	<220>																
	<221>	CDS															
20	<222>	(1)..(1500)															
	<400>	56															
	gtg	ata	gca	ata	tat	gta	agg	gta	tcg	acc	gag	gaa	caa	gcg	atc	aag	48
25	Val	Ile	Ala	Ile	Tyr	Val	Arg	Val	Ser	Thr	Glu	Glu	Gln	Ala	Ile	Lys	
	1				5					10					15		
	gga	tcg	agc	atc	gac	agc	caa	atc	gag	gcc	tgt	ata	aag	aaa	gca	ggg	96
	Gly	Ser	Ser	Ile	Asp	Ser	Gln	Ile	Glu	Ala	Cys	Ile	Lys	Lys	Ala	Gly	
				20					25					30			
30	act	aaa	gat	gtg	ctg	aag	tat	gca	gat	gaa	gga	ttt	tca	gga	gag	ctt	144
	Thr	Lys	Asp	Val	Leu	Lys	Tyr	Ala	Asp	Glu	Gly	Phe	Ser	Gly	Glu	Leu	
			35					40					45				
35	tta	gaa	cgt	ccg	gct	ttg	aat	cgc	ttg	agg	gag	gat	gca	agc	aag	gga	192
	Leu	Glu	Arg	Pro	Ala	Leu	Asn	Arg	Leu	Arg	Glu	Asp	Ala	Ser	Lys	Gly	
		50					55					60					
40	ctt	ata	agt	caa	gtc	att	tgt	tac	gat	cct	gac	cgt	ctt	tct	cgg	aaa	240
	Leu	Ile	Ser	Gln	Val	Ile	Cys	Tyr	Asp	Pro	Asp	Arg	Leu	Ser	Arg	Lys	
	65				70					75						80	
	tta	atg	aat	cag	cta	atc	att	gat	gac	gaa	ttg	cga	aag	cga	aac	ata	288
45	Leu	Met	Asn	Gln	Leu	Ile	Ile	Asp	Asp	Glu	Leu	Arg	Lys	Arg	Asn	Ile	
				85						90					95		
	cct	ttg	att	ttt	gta	aat	ggt	gaa	tac	gcc	aat	tct	cca	gaa	ggt	caa	336
	Pro	Leu	Ile	Phe	Val	Asn	Gly	Glu	Tyr	Ala	Asn	Ser	Pro	Glu	Gly	Gln	
				100					105					110			
50	ttg	ttt	ttc	gca	atg	cgc	ggg	gca	atc	tca	gaa	ttt	gaa	aaa	gcc	aaa	384
	Leu	Phe	Phe	Ala	Met	Arg	Gly	Ala	Ile	Ser	Glu	Phe	Glu	Lys	Ala	Lys	
			115				120					125					
55	atc	aaa	gaa	cgg	aca	tca	agc	ggc	cga	ctt	caa	aaa	atg	aaa	aaa	ggc	432
	Ile	Lys	Glu	Arg	Thr	Ser	Ser	Gly	Arg	Leu	Gln	Lys	Met	Lys	Lys	Gly	
		130					135					140					
60	atg	atc	att	aaa	gat	tct	aaa	cta	tat	ggc	tat	aaa	ttt	gtt	aaa	gag	480
	Met	Ile	Ile	Lys	Asp	Ser	Lys	Leu	Tyr	Gly	Tyr	Lys	Phe	Val	Lys	Glu	
	145					150					155					160	
	aaa	aga	act	ctt	gag	ata	tta	gaa	gag	gaa	gca	aaa	atc	att	cgg	atg	528
65	Lys	Arg	Thr	Leu	Glu	Ile	Leu	Glu	Glu	Glu	Ala	Lys	Ile	Ile	Arg	Met	
				165					170						175		
	att	ttt	aac	tat	ttc	acc	gat	cat	aaa	agc	cct	ttt	ttc	ggc	aga	gta	576

34

	Ile	Phe	Asn	Tyr	Phe	Thr	Asp	His	Lys	Ser	Pro	Phe	Phe	Gly	Arg	Val	
				180					185					190			
5	aat	ggt	att	gct	cta	cat	tta	act	cag	atg	ggg	gtt	aaa	aca	aaa	aaa	624
	Asn	Gly	Ile	Ala	Leu	His	Leu	Thr	Gln	Met	Gly	Val	Lys	Thr	Lys	Lys	
			195					200					205				
10	ggc	gcc	aaa	gta	tgg	cac	agg	cag	gtt	gtt	cgg	caa	ata	tta	atg	aac	672
	Gly	Ala	Lys	Val	Trp	His	Arg	Gln	Val	Val	Arg	Gln	Ile	Leu	Met	Asn	
			210				215					220					
15	tct	tcc	tat	aag	ggt	gaa	cat	aga	cag	tat	aaa	tat	gat	aca	gag	ggt	720
	Ser	Ser	Tyr	Lys	Gly	Glu	His	Arg	Gln	Tyr	Lys	Tyr	Asp	Thr	Glu	Gly	
	225					230					235					240	
20	tcc	tat	gtt	tca	aag	cag	gca	ggg	aac	aaa	tct	ata	att	aaa	ata	agg	768
	Ser	Tyr	Val	Ser	Lys	Gln	Ala	Gly	Asn	Lys	Ser	Ile	Ile	Lys	Ile	Arg	
					245					250					255		
25	cct	gaa	gaa	gaa	caa	atc	act	gtg	aca	att	cca	gca	att	gtt	cca	gct	816
	Pro	Glu	Glu	Glu	Gln	Ile	Thr	Val	Thr	Ile	Pro	Ala	Ile	Val	Pro	Ala	
				260					265					270			
30	gaa	caa	tgg	gat	tat	gct	caa	gaa	ctc	tta	ggt	caa	agt	aaa	aga	aaa	864
	Glu	Gln	Trp	Asp	Tyr	Ala	Gln	Glu	Leu	Leu	Gly	Gln	Ser	Lys	Arg	Lys	
			275					280					285				
35	cac	ttg	agt	atc	agc	cct	cac	aat	tac	ttg	tta	tcg	ggt	ttg	gtt	aga	912
	His	Leu	Ser	Ile	Ser	Pro	His	Asn	Tyr	Leu	Leu	Ser	Gly	Leu	Val	Arg	
		290					295					300					
40	tgc	gga	aaa	tgc	gga	aat	acc	atg	aca	ggg	aag	aaa	aga	aaa	tca	cat	960
	Cys	Gly	Lys	Cys	Gly	Asn	Thr	Met	Thr	Gly	Lys	Lys	Arg	Lys	Ser	His	
	305					310					315					320	
45	ggt	aaa	gac	tac	tat	gta	tat	act	tgc	cgg	aaa	aat	tat	tct	ggc	gca	1008
	Gly	Lys	Asp	Tyr	Tyr	Val	Tyr	Thr	Cys	Arg	Lys	Asn	Tyr	Ser	Gly	Ala	
					325					330					335		
50	aag	gac	cgc	ggc	tgc	gga	aaa	gaa	atg	tct	gag	aat	aaa	ttg	aac	cgg	1056
	Lys	Asp	Arg	Gly	Cys	Gly	Lys	Glu	Met	Ser	Glu	Asn	Lys	Leu	Asn	Arg	
				340					345					350			
55	cat	gta	tgg	ggt	gaa	att	ttt	aaa	ttc	atc	aca	aat	cct	caa	aag	tat	1104
	His	Val	Trp	Gly	Glu	Ile	Phe	Lys	Phe	Ile	Thr	Asn	Pro	Gln	Lys	Tyr	
			355					360					365				
60	gtt	tct	ttt	aaa	gag	gct	gaa	caa	tca	aat	cac	ctg	tct	gat	gaa	tta	1152
	Val	Ser	Phe	Lys	Glu	Ala	Glu	Gln	Ser	Asn	His	Leu	Ser	Asp	Glu	Leu	
		370					375					380					
65	gaa	ctt	att	gaa	aaa	gag	ata	gag	aaa	aca	aaa	aaa	ggc	cgc	aag	cgt	1200
	Glu	Leu	Ile	Glu	Lys	Glu	Ile	Glu	Lys	Thr	Lys	Lys	Gly	Arg	Lys	Arg	
	385					390					395					400	
70	ctt	tta	acg	cta	atc	agc	cta	agc	gat	gac	gat	gat	tta	gac	ata	gat	1248
	Leu	Leu	Thr	Leu	Ile	Ser	Leu	Ser	Asp	Asp	Asp	Asp	Leu	Asp	Ile	Asp	
					405					410					415		
75	gaa	atc	aaa	gca	caa	att	att	gaa	ctg	caa	aaa	aag	caa	aat	cag	ctt	1296
	Glu	Ile	Lys	Ala	Gln	Ile	Ile	Glu	Leu	Gln	Lys	Lys	Gln	Asn	Gln	Leu	
				420					425					430			
80	act	gaa	aag	tgt	aac	aga	atc	cag	tca	aaa	atg	aaa	gtc	cta	gat	gat	1344
	Thr	Glu	Lys	Cys	Asn	Arg	Ile	Gln	Ser	Lys	Met	Lys	Val	Leu	Asp	Asp	
			435					440					445				

35

	acg	agc	tca	agt	gaa	aat	gct	cta	aaa	aga	gcc	atc	gac	tat	ttt	caa	1392
	Thr	Ser	Ser	Ser	Glu	Asn	Ala	Leu	Lys	Arg	Ala	Ile	Asp	Tyr	Phe	Gln	
		450					455					460					
5	tca	atc	ggg	gca	gat	aac	tta	act	ctt	gaa	gat	aaa	aaa	aca	att	gtt	1440
	Ser	Ile	Gly	Ala	Asp	Asn	Leu	Thr	Leu	Glu	Asp	Lys	Lys	Thr	Ile	Val	
	465					470					475					480	
10	aac	ttt	atc	gtg	aaa	gaa	gtt	acc	att	gtg	gat	tct	gac	acc	ata	tat	1488
	Asn	Phe	Ile	Val	Lys	Glu	Val	Thr	Ile	Val	Asp	Ser	Asp	Thr	Ile	Tyr	
					485					490					495		
15	att	gaa	acg	tat	taa												1503
	Ile	Glu	Thr	Tyr													
				500													
	<210> 57																
	<211> 500																
20	<212> PRT																
	<213> Cisa recombinase																
	<400> 57																
25	Val	Ile	Ala	Ile	Tyr	Val	Arg	Val	Ser	Thr	Glu	Glu	Gln	Ala	Ile	Lys	
	1				5					10					15		
	Gly	Ser	Ser	Ile	Asp	Ser	Gln	Ile	Glu	Ala	Cys	Ile	Lys	Lys	Ala	Gly	
				20					25					30			
30	Thr	Lys	Asp	Val	Leu	Lys	Tyr	Ala	Asp	Glu	Gly	Phe	Ser	Gly	Glu	Leu	
			35					40					45				
	Leu	Glu	Arg	Pro	Ala	Leu	Asn	Arg	Leu	Arg	Glu	Asp	Ala	Ser	Lys	Gly	
	50						55					60					
35	Leu	Ile	Ser	Gln	Val	Ile	Cys	Tyr	Asp	Pro	Asp	Arg	Leu	Ser	Arg	Lys	
	65					70				75						80	
40	Leu	Met	Asn	Gln	Leu	Ile	Ile	Asp	Asp	Glu	Leu	Arg	Lys	Arg	Asn	Ile	
				85						90					95		
	Pro	Leu	Ile	Phe	Val	Asn	Gly	Glu	Tyr	Ala	Asn	Ser	Pro	Glu	Gly	Gln	
				100					105					110			
45	Leu	Phe	Phe	Ala	Met	Arg	Gly	Ala	Ile	Ser	Glu	Phe	Glu	Lys	Ala	Lys	
		115					120						125				
	Ile	Lys	Glu	Arg	Thr	Ser	Ser	Gly	Arg	Leu	Gln	Lys	Met	Lys	Lys	Gly	
	130						135					140					
50	Met	Ile	Ile	Lys	Asp	Ser	Lys	Leu	Tyr	Gly	Tyr	Lys	Phe	Val	Lys	Glu	
	145					150					155					160	
55	Lys	Arg	Thr	Leu	Glu	Ile	Leu	Glu	Glu	Glu	Ala	Lys	Ile	Ile	Arg	Met	
				165					170						175		
	Ile	Phe	Asn	Tyr	Phe	Thr	Asp	His	Lys	Ser	Pro	Phe	Phe	Gly	Arg	Val	
			180					185						190			
60	Asn	Gly	Ile	Ala	Leu	His	Leu	Thr	Gln	Met	Gly	Val	Lys	Thr	Lys	Lys	
		195					200						205				
	Gly	Ala	Lys	Val	Trp	His	Arg	Gln	Val	Val	Arg	Gln	Ile	Leu	Met	Asn	
	210					215					220						
65	Ser	Ser	Tyr	Lys	Gly	Glu	His	Arg	Gln	Tyr	Lys	Tyr	Asp	Thr	Glu	Gly	
	225					230				235						240	

36

Ser Tyr Val Ser Lys Gln Ala Gly Asn Lys Ser Ile Ile Lys Ile Arg
 245 250 255
 5 Pro Glu Glu Glu Gln Ile Thr Val Thr Ile Pro Ala Ile Val Pro Ala
 260 265 270
 Glu Gln Trp Asp Tyr Ala Gln Glu Leu Leu Gly Gln Ser Lys Arg Lys
 275 280 285
 10 His Leu Ser Ile Ser Pro His Asn Tyr Leu Leu Ser Gly Leu Val Arg
 290 295 300
 Cys Gly Lys Cys Gly Asn Thr Met Thr Gly Lys Lys Arg Lys Ser His
 305 310 315 320
 Gly Lys Asp Tyr Tyr Val Tyr Thr Cys Arg Lys Asn Tyr Ser Gly Ala
 325 330 335
 20 Lys Asp Arg Gly Cys Gly Lys Glu Met Ser Glu Asn Lys Leu Asn Arg
 340 345 350
 His Val Trp Gly Glu Ile Phe Lys Phe Ile Thr Asn Pro Gln Lys Tyr
 355 360 365
 25 Val Ser Phe Lys Glu Ala Glu Gln Ser Asn His Leu Ser Asp Glu Leu
 370 375 380
 Glu Leu Ile Glu Lys Glu Ile Glu Lys Thr Lys Lys Gly Arg Lys Arg
 385 390 395 400
 Leu Leu Thr Leu Ile Ser Leu Ser Asp Asp Asp Asp Leu Asp Ile Asp
 405 410 415
 35 Glu Ile Lys Ala Gln Ile Ile Glu Leu Gln Lys Lys Gln Asn Gln Leu
 420 425 430
 Thr Glu Lys Cys Asn Arg Ile Gln Ser Lys Met Lys Val Leu Asp Asp
 435 440 445
 40 Thr Ser Ser Ser Glu Asn Ala Leu Lys Arg Ala Ile Asp Tyr Phe Gln
 450 455 460
 Ser Ile Gly Ala Asp Asn Leu Thr Leu Glu Asp Lys Lys Thr Ile Val
 465 470 475 480
 Asn Phe Ile Val Lys Glu Val Thr Ile Val Asp Ser Asp Thr Ile Tyr
 485 490 495
 50 Ile Glu Thr Tyr
 500
 55 <210> 58
 <211> 1545
 <212> DNA
 <213> XisF recombinase
 60 <220>
 <221> CDS
 <222> (1)..(1542)
 <400> 58
 65 atg gaa aat tgg ggt tac gcg aga gtg agc ggt gag gaa cag caa aca 48
 Met Glu Asn Trp Gly Tyr Ala Arg Val Ser Gly Glu Glu Gln Gln Thr
 1 5 10 15

5	gat	aaa	ggt	gcg	ttg	cgt	aaa	caa	ata	gaa	cgc	ttg	cgt	aat	gct	gga	96
	Asp	Lys	Gly	Ala	Leu	Arg	Lys	Gln	Ile	Glu	Arg	Leu	Arg	Asn	Ala	Gly	
				20					25					30			
10	tgt	tca	aaa	gtg	tac	tgg	gat	att	caa	tcg	cgg	aca	act	gaa	gtc	aga	144
	Cys	Ser	Lys	Val	Tyr	Trp	Asp	Ile	Gln	Ser	Arg	Thr	Thr	Glu	Val	Arg	
			35					40					45				
15	gaa	ggg	cta	caa	caa	tta	att	aat	gac	tta	aag	aca	tct	tca	aca	ggt	192
	Glu	Gly	Leu	Gln	Gln	Leu	Ile	Asn	Asp	Leu	Lys	Thr	Ser	Ser	Thr	Gly	
		50					55				60						
20	aag	gta	aaa	tca	ctg	caa	ttt	acc	cgc	att	gat	cgc	atc	ggc	tca	tca	240
	Lys	Val	Lys	Ser	Leu	Gln	Phe	Thr	Arg	Ile	Asp	Arg	Ile	Gly	Ser	Ser	
	65				70					75						80	
25	tcg	cgg	ttg	ttt	tat	tca	ttg	tta	gag	gta	tta	cgt	tcc	aag	gga	att	288
	Ser	Arg	Leu	Phe	Tyr	Ser	Leu	Leu	Glu	Val	Leu	Arg	Ser	Lys	Gly	Ile	
				85					90						95		
30	aaa	ctg	ata	gcc	tta	gat	caa	ggc	gtt	gac	cca	gac	agc	ctt	ggc	ggg	336
	Lys	Leu	Ile	Ala	Leu	Asp	Gln	Gly	Val	Asp	Pro	Asp	Ser	Leu	Gly	Gly	
			100					105						110			
35	gaa	cta	aca	att	gat	atg	tta	ctg	gct	gct	gcc	aaa	ttt	gag	gta	aga	384
	Glu	Leu	Thr	Ile	Asp	Met	Leu	Leu	Ala	Ala	Ala	Lys	Phe	Glu	Val	Arg	
			115				120						125				
40	atg	gtg	acg	gag	agg	tta	aaa	agc	gaa	cgt	cgt	cat	agg	gtg	aac	caa	432
	Met	Val	Thr	Glu	Arg	Leu	Lys	Ser	Glu	Arg	Arg	His	Arg	Val	Asn	Gln	
		130				135						140					
45	gga	aaa	agt	cac	cga	gtt	gcc	cca	tta	gga	tac	cgc	aaa	gat	aaa	gat	480
	Gly	Lys	Ser	His	Arg	Val	Ala	Pro	Leu	Gly	Tyr	Arg	Lys	Asp	Lys	Asp	
	145				150					155				160			
50	aaa	tat	ata	cgc	gat	cgc	tca	cca	tgt	gtt	tgc	tta	cta	gaa	gga	cgc	528
	Lys	Tyr	Ile	Arg	Asp	Arg	Ser	Pro	Cys	Val	Cys	Leu	Leu	Glu	Gly	Arg	
				165					170						175		
55	aga	gaa	tta	acg	gtg	tct	gac	tta	gcc	cag	tat	att	ttt	cac	act	ttt	576
	Arg	Glu	Leu	Thr	Val	Ser	Asp	Leu	Ala	Gln	Tyr	Ile	Phe	His	Thr	Phe	
			180					185					190				
60	ttt	gag	tgc	ggt	tcc	gtt	gct	gct	act	gtg	cgt	aag	ctg	cac	tca	gat	624
	Phe	Glu	Cys	Gly	Ser	Val	Ala	Ala	Thr	Val	Arg	Lys	Leu	His	Ser	Asp	
		195				200						205					
65	ttt	ggt	ata	gaa	aca	aaa	gtt	ctg	aat	tgg	aac	aag	cta	gaa	aaa	tct	672
	Phe	Gly	Ile	Glu	Thr	Lys	Val	Leu	Asn	Trp	Asn	Lys	Leu	Glu	Lys	Ser	
		210				215						220					
70	tcc	cgg	att	gtt	ggc	gac	gac	gac	tta	gat	aaa	att	gca	ttt	aca	cca	720
	Ser	Arg	Ile	Val	Gly	Asp	Asp	Asp	Leu	Asp	Lys	Ile	Ala	Phe	Thr	Pro	
	225				230					235						240	
75	aat	aaa	act	aac	cac	ccc	ttg	cgt	tat	ccc	tgg	tct	ggg	cta	aga	tgg	768
	Asn	Lys	Thr	Asn	His	Pro	Leu	Arg	Tyr	Pro	Trp	Ser	Gly	Leu	Arg	Trp	
				245					250						255		
80	tca	atc	cct	ggt	tta	aaa	gcg	tta	tta	gtt	aac	cct	gtt	tac	gcc	ggg	816
	Ser	Ile	Pro	Gly	Leu	Lys	Ala	Leu	Leu	Val	Asn	Pro	Val	Tyr	Ala	Gly	
			260					265					270				
85	ggt	ttg	ccc	ttt	gat	act	tac	gtt	aaa	tca	aaa	gga	aaa	cgc	aag	cat	864
	Gly	Leu	Pro	Phe	Asp	Thr	Tyr	Val	Lys	Ser	Lys	Gly	Lys	Arg	Lys	His	

38

	275	280	285	
5	ttt gac gag tgg aaa gta Phe Asp Glu Trp Lys Val 290	aaa tgg gga acc cac Lys Trp Gly Thr His 295	gac gat gag gca atc Asp Asp Glu Ala Ile 300	912
10	att acc tgt gag gaa cat Ile Thr Cys Glu Glu His 305	gaa aga ata aaa cag Glu Arg Ile Lys Gln 310	atg att cga gac aat Met Ile Arg Asp Asn 315	960
	cgc aat aat cga tgg gct Arg Asn Asn Arg Trp Ala 325	gca aga gaa gaa aac Ala Arg Glu Glu Asn 330	gaa gta aac cca ttt Glu Val Asn Pro Phe 335	1008
15	tct aat tta ctt aaa tgt Ser Asn Leu Leu Lys Cys 340	acc cat tgc ggc ggc Thr His Cys Gly Gly 345	tca atg aca cgc cac Ser Met Thr Arg His 350	1056
20	gcc aaa cgt gta gat aag Ala Lys Arg Val Asp Lys 355	agt gga caa gct atc Ser Gly Gln Ala Ile 360	tat tat tat cag tgc Tyr Tyr Tyr Gln Cys 365	1104
25	cga ttg tat aaa gct ggc Arg Leu Tyr Lys Ala Gly 370	aac tgt agc aat aaa Asn Cys Ser Asn Lys 375	aat atg att tca tcc Asn Met Ile Ser Ser 380	1152
30	aaa ata tta gat atc caa Lys Ile Leu Asp Ile Gln 385	gta atg gat tta ttg Val Met Asp Leu Leu 390	gca caa gaa gcc gaa Ala Gln Glu Ala Glu 395	1200
	cgt tta gca aat ttg gtg Arg Leu Ala Asn Leu Val 405	gaa aca gat gag ccg Glu Thr Asp Glu Pro 410	ctt att gta gaa gaa Leu Ile Val Glu Glu 415	1248
35	ccc cca gaa gta aaa acg Pro Pro Glu Val Lys Thr 420	ctg cgc gca tcc ctg Leu Arg Ala Ser Leu 425	aat agt ctg gaa aca Asn Ser Ser Glu Thr 430	1296
40	ttg cca gca agt tca gca Leu Pro Ala Ser Ser Ala 435	att gaa caa att aaa Ile Glu Gln Ile Lys 440	aat gac ctc aaa gaa Asn Asp Leu Lys Glu 445	1344
45	cag att gcg atc gca cta Gln Ile Ala Ile Ala Leu 450	gga gca acc aat aat Gly Ala Thr Asn Asn 455	gct tct aaa caa tct Ala Ser Lys Gln Ser 460	1392
50	ctg att gcc aag gaa aga Leu Ile Ala Lys Glu Arg 465	att ata caa gct ttt Ile Ile Gln Ala Phe 470	gct cat aaa agt tac Ala His Lys Ser Tyr 475	1440
	tgg caa gga cta aac gct Trp Gln Gly Leu Asn Ala 485	caa gat aaa cga gca Gln Asp Lys Arg Ala 490	atc ctc aat ggt tgc Ile Leu Asn Gly Cys 495	1488
55	gta aaa aaa atc tcc gta Val Lys Lys Ile Ser Val 500	gat ggt aac ttt gtt Asp Gly Asn Phe Val 505	aca gct att gag tat Thr Ala Ile Glu Tyr 510	1536
60	cgt tac tag Arg Tyr			1545
65	<210> 59 <211> 514 <212> PRT <213> XisF recombinase			

39

<400> 59

	Met	Glu	Asn	Trp	Gly	Tyr	Ala	Arg	Val	Ser	Gly	Glu	Glu	Gln	Gln	Thr	
	1				5					10					15		
5	Asp	Lys	Gly	Ala	Leu	Arg	Lys	Gln	Ile	Glu	Arg	Leu	Arg	Asn	Ala	Gly	
				20					25					30			
	Cys	Ser	Lys	Val	Tyr	Trp	Asp	Ile	Gln	Ser	Arg	Thr	Thr	Glu	Val	Arg	
			35					40					45				
10	Glu	Gly	Leu	Gln	Gln	Leu	Ile	Asn	Asp	Leu	Lys	Thr	Ser	Ser	Thr	Gly	
		50					55					60					
	Lys	Val	Lys	Ser	Leu	Gln	Phe	Thr	Arg	Ile	Asp	Arg	Ile	Gly	Ser	Ser	
15	65					70					75					80	
	Ser	Arg	Leu	Phe	Tyr	Ser	Leu	Leu	Glu	Val	Leu	Arg	Ser	Lys	Gly	Ile	
				85						90					95		
20	Lys	Leu	Ile	Ala	Leu	Asp	Gln	Gly	Val	Asp	Pro	Asp	Ser	Leu	Gly	Gly	
				100					105					110			
	Glu	Leu	Thr	Ile	Asp	Met	Leu	Leu	Ala	Ala	Ala	Lys	Phe	Glu	Val	Arg	
25			115					120					125				
	Met	Val	Thr	Glu	Arg	Leu	Lys	Ser	Glu	Arg	Arg	His	Arg	Val	Asn	Gln	
	130						135					140					
	Gly	Lys	Ser	His	Arg	Val	Ala	Pro	Leu	Gly	Tyr	Arg	Lys	Asp	Lys	Asp	
30	145					150					155					160	
	Lys	Tyr	Ile	Arg	Asp	Arg	Ser	Pro	Cys	Val	Cys	Leu	Leu	Glu	Gly	Arg	
				165						170					175		
35	Arg	Glu	Leu	Thr	Val	Ser	Asp	Leu	Ala	Gln	Tyr	Ile	Phe	His	Thr	Phe	
				180					185					190			
	Phe	Glu	Cys	Gly	Ser	Val	Ala	Ala	Thr	Val	Arg	Lys	Leu	His	Ser	Asp	
			195					200					205				
40	Phe	Gly	Ile	Glu	Thr	Lys	Val	Leu	Asn	Trp	Asn	Lys	Leu	Glu	Lys	Ser	
	210						215					220					
	Ser	Arg	Ile	Val	Gly	Asp	Asp	Asp	Leu	Asp	Lys	Ile	Ala	Phe	Thr	Pro	
45	225					230					235					240	
	Asn	Lys	Thr	Asn	His	Pro	Leu	Arg	Tyr	Pro	Trp	Ser	Gly	Leu	Arg	Trp	
				245						250					255		
50	Ser	Ile	Pro	Gly	Leu	Lys	Ala	Leu	Leu	Val	Asn	Pro	Val	Tyr	Ala	Gly	
				260					265					270			
	Gly	Leu	Pro	Phe	Asp	Thr	Tyr	Val	Lys	Ser	Lys	Gly	Lys	Arg	Lys	His	
			275					280					285				
55	Phe	Asp	Glu	Trp	Lys	Val	Lys	Trp	Gly	Thr	His	Asp	Asp	Glu	Ala	Ile	
	290						295					300					
	Ile	Thr	Cys	Glu	Glu	His	Glu	Arg	Ile	Lys	Gln	Met	Ile	Arg	Asp	Asn	
60	305					310					315					320	
	Arg	Asn	Asn	Arg	Trp	Ala	Ala	Arg	Glu	Glu	Asn	Glu	Val	Asn	Pro	Phe	
				325						330					335		
65	Ser	Asn	Leu	Leu	Lys	Cys	Thr	His	Cys	Gly	Gly	Ser	Met	Thr	Arg	His	
			340						345						350		

40

	Ala	Lys	Arg	Val	Asp	Lys	Ser	Gly	Gln	Ala	Ile	Tyr	Tyr	Tyr	Gln	Cys	
			355					360					365				
5	Arg	Leu	Tyr	Lys	Ala	Gly	Asn	Cys	Ser	Asn	Lys	Asn	Met	Ile	Ser	Ser	
		370					375					380					
	Lys	Ile	Leu	Asp	Ile	Gln	Val	Met	Asp	Leu	Leu	Ala	Gln	Glu	Ala	Glu	
	385					390					395					400	
10	Arg	Leu	Ala	Asn	Leu	Val	Glu	Thr	Asp	Glu	Pro	Leu	Ile	Val	Glu	Glu	
				405						410					415		
	Pro	Pro	Glu	Val	Lys	Thr	Leu	Arg	Ala	Ser	Leu	Asn	Ser	Leu	Glu	Thr	
				420					425					430			
15	Leu	Pro	Ala	Ser	Ser	Ala	Ile	Glu	Gln	Ile	Lys	Asn	Asp	Leu	Lys	Glu	
			435					440					445				
20	Gln	Ile	Ala	Ile	Ala	Leu	Gly	Ala	Thr	Asn	Asn	Ala	Ser	Lys	Gln	Ser	
	450						455					460					
	Leu	Ile	Ala	Lys	Glu	Arg	Ile	Ile	Gln	Ala	Phe	Ala	His	Lys	Ser	Tyr	
	465					470					475					480	
25	Trp	Gln	Gly	Leu	Asn	Ala	Gln	Asp	Lys	Arg	Ala	Ile	Leu	Asn	Gly	Cys	
				485						490					495		
	Val	Lys	Lys	Ile	Ser	Val	Asp	Gly	Asn	Phe	Val	Thr	Ala	Ile	Glu	Tyr	
				500					505					510			
30	Arg	Tyr															
35																	
	<210>	60															
	<211>	2124															
	<212>	DNA															
	<213>	Transposon Tn4451															
40																	
	<220>																
	<221>	CDS															
	<222>	(1)..(2121)															
45	<400>	60															
	atg	tca	agg	act	tca	aga	att	aca	gca	ctt	tac	gag	cgt	ttg	tca	aga	48
	Met	Ser	Arg	Thr	Ser	Arg	Ile	Thr	Ala	Leu	Tyr	Glu	Arg	Leu	Ser	Arg	
	1				5					10				15			
50	gat	gat	gac	ctt	act	ggc	gag	agt	aat	tct	att	acc	aat	caa	aag	aaa	96
	Asp	Asp	Asp	Leu	Thr	Gly	Glu	Ser	Asn	Ser	Ile	Thr	Asn	Gln	Lys	Lys	
				20					25					30			
55	tac	ctc	gaa	gat	tat	gcc	cgt	agg	aat	ggt	ttt	gag	aac	att	cgc	cat	144
	Tyr	Leu	Glu	Asp	Tyr	Ala	Arg	Arg	Asn	Gly	Phe	Glu	Asn	Ile	Arg	His	
			35					40					45				
60	ttt	acc	gat	gac	gga	ttt	tog	ggt	gta	aat	ttc	aat	cgc	cct	ggc	ttt	192
	Phe	Thr	Asp	Asp	Gly	Phe	Ser	Gly	Val	Asn	Phe	Asn	Arg	Pro	Gly	Phe	
			50				55					60					
	caa	tct	ctg	ata	aaa	gaa	gtt	gaa	gca	gga	aat	gta	gaa	acc	ttg	att	240
	Gln	Ser	Leu	Ile	Lys	Glu	Val	Glu	Ala	Gly	Asn	Val	Glu	Thr	Leu	Ile	
	65					70				75						80	
65	gtt	aag	gat	atg	agc	cga	ttg	ggg	cga	aat	tat	ctg	caa	gta	ggg	ttt	288
	Val	Lys	Asp	Met	Ser	Arg	Leu	Gly	Arg	Asn	Tyr	Leu	Gln	Val	Gly	Phe	

	85								41	95							
									90								
5	tat	acg	gaa	gtt	ctg	ttt	cca	cag	aaa	aat	gtc	cgt	ttc	ctt	gca	att	336
	Tyr	Thr	Glu	Val	Leu	Phe	Pro	Gln	Lys	Asn	Val	Arg	Phe	Leu	Ala	Ile	
				100					105					110			
10	aac	aac	agt	att	gac	agt	aac	aac	gct	tcg	gat	aat	gac	ttt	gct	cog	384
	Asn	Asn	Ser	Ile	Asp	Ser	Asn	Asn	Ala	Ser	Asp	Asn	Asp	Phe	Ala	Pro	
			115					120					125				
15	ttt	ttg	aat	att	atg	aac	gaa	tgg	tat	gcc	aaa	gac	aca	agc	aac	aaa	432
	Phe	Leu	Asn	Ile	Met	Asn	Glu	Trp	Tyr	Ala	Lys	Asp	Thr	Ser	Asn	Lys	
			130				135					140					
20	atc	aag	gct	ata	ttc	gat	gcc	cgt	atg	aaa	gac	gga	aag	cgt	tgt	agc	480
	Ile	Lys	Ala	Ile	Phe	Asp	Ala	Arg	Met	Lys	Asp	Gly	Lys	Arg	Cys	Ser	
	145					150					155					160	
25	ggt	tct	atc	cct	tat	ggg	tat	aac	cga	ctg	ccg	agc	gac	aaa	caa	acg	528
	Gly	Ser	Ile	Pro	Tyr	Gly	Tyr	Asn	Arg	Leu	Pro	Ser	Asp	Lys	Gln	Thr	
					165					170					175		
30	ctt	gtg	gtt	gac	cct	gtg	gct	tcg	gaa	gtg	gta	aag	cgt	atc	ttt	act	576
	Leu	Val	Val	Asp	Pro	Val	Ala	Ser	Glu	Val	Val	Lys	Arg	Ile	Phe	Thr	
				180					185					190			
35	ctt	gcc	aat	gat	ggc	aaa	agt	aca	agg	gca	atc	gca	gaa	ata	ctg	acc	624
	Leu	Ala	Asn	Asp	Gly	Lys	Ser	Thr	Arg	Ala	Ile	Ala	Glu	Ile	Leu	Thr	
			195					200					205				
40	gaa	gaa	aaa	gtt	tta	acc	cct	gcg	gca	tac	gca	aag	gaa	tac	cac	ccc	672
	Glu	Glu	Lys	Val	Leu	Thr	Pro	Ala	Ala	Tyr	Ala	Lys	Glu	Tyr	His	Pro	
			210				215					220					
45	gaa	cag	tac	aac	ggc	aac	aag	ttc	aca	aac	cct	tat	ctt	tgg	gca	atg	720
	Glu	Gln	Tyr	Asn	Gly	Asn	Lys	Phe	Thr	Asn	Pro	Tyr	Leu	Trp	Ala	Met	
	225				230					235						240	
50	tca	acg	ata	aga	aat	att	tta	ggc	agg	cag	gaa	tat	ctc	ggt	cac	acc	768
	Ser	Thr	Ile	Arg	Asn	Ile	Leu	Gly	Arg	Gln	Glu	Tyr	Leu	Gly	His	Thr	
					245					250					255		
55	gtt	ttg	cga	aag	tcg	gta	agc	aca	aat	ttc	aaa	ctt	cac	aag	aga	aaa	816
	Val	Leu	Arg	Lys	Ser	Val	Ser	Thr	Asn	Phe	Lys	Leu	His	Lys	Arg	Lys	
				260					265					270			
60	agc	aca	gac	gaa	gaa	gaa	cag	tat	gta	ttt	ccg	aat	aca	cac	gag	cct	864
	Ser	Thr	Asp	Glu	Glu	Glu	Gln	Tyr	Val	Phe	Pro	Asn	Thr	His	Glu	Pro	
			275					280					285				
65	atc	ata	tcg	cag	gaa	ctt	tgg	gac	agc	gtt	caa	aaa	cgc	aga	agc	aga	912
	Ile	Ile	Ser	Gln	Glu	Leu	Trp	Asp	Ser	Val	Gln	Lys	Arg	Arg	Ser	Arg	
			290				295					300					
70	gta	aat	cgt	gcc	tcg	gct	tgg	gga	acg	cac	agc	aac	cgt	tta	agc	gga	960
	Val	Asn	Arg	Ala	Ser	Ala	Trp	Gly	Thr	His	Ser	Asn	Arg	Leu	Ser	Gly	
	305					310					315					320	
75	tat	ttg	tac	tgt	gcc	gat	tgc	gga	aga	aga	atg	act	ttg	cag	aca	cat	1008
	Tyr	Leu	Tyr	Cys	Ala	Asp	Cys	Gly	Arg	Arg	Met	Thr	Leu	Gln	Thr	His	
					325					330					335		
80	tac	agc	aaa	aaa	gac	ggg	tct	gtg	cag	tat	tct	tac	cgt	tgc	ggg	ggg	1056
	Tyr	Ser	Lys	Lys	Asp	Gly	Ser	Val	Gln	Tyr	Ser	Tyr	Arg	Cys	Gly	Gly	
				340					345					350			
85	tat	gca	agc	aga	gtg	aac	agt	tgt	acc	agt	cat	tcg	att	agt	acc	gat	1104

42

	Tyr	Ala	Ser	Arg	Val	Asn	Ser	Cys	Thr	Ser	His	Ser	Ile	Ser	Thr	Asp	
			355					360					365				
5	aat	ggt	gaa	gcc	ttg	ata	tta	tca	tct	gtc	aaa	cgc	ttt	tca	agg	ttt	1152
	Asn	Val	Glu	Ala	Leu	Ile	Leu	Ser	Ser	Val	Lys	Arg	Phe	Ser	Arg	Phe	
		370					375					380					
10	ggt	ctg	aat	gat	gaa	caa	gca	ttt	gct	ttg	gaa	ctg	caa	tct	ctt	tgg	1200
	Val	Leu	Asn	Asp	Glu	Gln	Ala	Phe	Ala	Leu	Glu	Leu	Gln	Ser	Leu	Trp	
	385					390					395					400	
15	aat	gaa	aaa	cag	gag	gaa	aag	ccg	aaa	cac	aat	caa	tcg	gaa	ctg	caa	1248
	Asn	Glu	Lys	Gln	Glu	Glu	Lys	Pro	Lys	His	Asn	Gln	Ser	Glu	Leu	Gln	
					405					410						415	
	cgc	tgt	cag	aaa	cgc	tat	gac	gaa	ctc	tct	acc	ctt	ggt	cgt	ggc	ttg	1296
	Arg	Cys	Gln	Lys	Arg	Tyr	Asp	Glu	Leu	Ser	Thr	Leu	Val	Arg	Gly	Leu	
				420					425					430			
20	tat	gaa	aat	ctt	atg	tcg	gga	tta	ctg	ccc	gaa	aga	cag	tat	aag	caa	1344
	Tyr	Glu	Asn	Leu	Met	Ser	Gly	Leu	Leu	Pro	Glu	Arg	Gln	Tyr	Lys	Gln	
			435					440					445				
25	ctg	atg	aaa	cag	tat	gat	gac	gag	cag	gca	gag	ttg	gaa	acg	aaa	atg	1392
	Leu	Met	Lys	Gln	Tyr	Asp	Asp	Glu	Gln	Ala	Glu	Leu	Glu	Thr	Lys	Met	
		450					455					460					
30	gaa	acg	atg	aaa	aca	gaa	ctt	gcc	gaa	gaa	aaa	gta	agt	tcc	ggt	gat	1440
	Glu	Thr	Met	Lys	Thr	Glu	Leu	Ala	Glu	Glu	Lys	Val	Ser	Ser	Val	Asp	
	465					470					475					480	
35	att	aag	cat	ttc	att	tcg	ctg	ata	cgc	aag	tgt	aaa	aat	cct	acg	gaa	1488
	Ile	Lys	His	Phe	Ile	Ser	Leu	Ile	Arg	Lys	Cys	Lys	Asn	Pro	Thr	Glu	
				485						490					495		
	atc	tcc	gat	aca	atg	ttt	aat	gaa	ctt	ggt	gat	aag	ata	gtg	ggt	tat	1536
	Ile	Ser	Asp	Thr	Met	Phe	Asn	Glu	Leu	Val	Asp	Lys	Ile	Val	Val	Tyr	
				500					505					510			
40	gaa	gca	gag	ggt	gtg	gga	aaa	gca	cga	aca	caa	aag	gtc	gat	att	tat	1584
	Glu	Ala	Glu	Gly	Val	Gly	Lys	Ala	Arg	Thr	Gln	Lys	Val	Asp	Ile	Tyr	
			515					520					525				
45	ttt	aac	tat	gtc	ggt	caa	gtg	gat	att	gcc	tat	acc	gaa	gaa	gaa	ctt	1632
	Phe	Asn	Tyr	Val	Gly	Gln	Val	Asp	Ile	Ala	Tyr	Thr	Glu	Glu	Glu	Leu	
		530					535					540					
50	gcc	gag	ata	gaa	aca	cag	aaa	gag	cag	gag	gaa	cag	caa	cgc	ttg	gca	1680
	Ala	Glu	Ile	Glu	Thr	Gln	Lys	Glu	Gln	Glu	Glu	Gln	Gln	Arg	Leu	Ala	
	545					550					555					560	
55	aga	cag	cgc	aag	cgt	gaa	aaa	gcc	tac	cga	gaa	aag	cga	aag	gca	cag	1728
	Arg	Gln	Arg	Lys	Arg	Glu	Lys	Ala	Tyr	Arg	Glu	Lys	Arg	Lys	Ala	Gln	
					565					570					575		
	aaa	atc	gct	gaa	aac	ggt	ggc	gaa	atc	ggt	aag	aca	aag	ggt	tgc	cct	1776
	Lys	Ile	Ala	Glu	Asn	Gly	Gly	Glu	Ile	Val	Lys	Thr	Lys	Val	Cys	Pro	
				580					585					590			
60	cat	tgc	aac	aaa	gag	ttt	atc	ccg	aca	agc	aac	cga	cag	gtg	ttc	tgt	1824
	His	Cys	Asn	Lys	Glu	Phe	Ile	Pro	Thr	Ser	Asn	Arg	Gln	Val	Phe	Cys	
			595					600					605				
65	tcc	aaa	gag	tgc	tgc	tat	caa	gca	agg	caa	gac	aaa	aag	aaa	aca	gac	1872
	Ser	Lys	Glu	Cys	Cys	Tyr	Gln	Ala	Arg	Gln	Asp	Lys	Lys	Lys	Thr	Asp	
		610					615					620					

43

	cga	gaa	gca	gaa	cga	gga	aat	cac	tat	tac	cga	cag	cgt	gta	tgt	gct	1920
	Arg	Glu	Ala	Glu	Arg	Gly	Asn	His	Tyr	Tyr	Arg	Gln	Arg	Val	Cys	Ala	
	625					630					635					640	
5	gtg	tgc	ggc	aat	tcc	tat	tgg	cct	aca	cac	agc	caa	cag	aaa	ttc	tgc	1968
	Val	Cys	Gly	Asn	Ser	Tyr	Trp	Pro	Thr	His	Ser	Gln	Gln	Lys	Phe	Cys	
					645					650					655		
10	tcc	gaa	gaa	tgt	caa	agg	gta	aat	cac	aat	aag	aaa	aca	ttg	gaa	ttt	2016
	Ser	Glu	Glu	Cys	Gln	Arg	Val	Asn	His	Asn	Lys	Lys	Thr	Leu	Glu	Phe	
				660					665					670			
15	tac	cac	cat	aaa	aaa	gaa	aag	gag	aag	ctg	caa	tgc	aaa	gat	tta	tca	2064
	Tyr	His	His	Lys	Lys	Glu	Lys	Glu	Lys	Leu	Gln	Cys	Lys	Asp	Leu	Ser	
			675					680					685				
20	cag	acg	aaa	gaa	cgg	gta	tcc	gat	atg	aac	tta	tcg	ggg	act	att	act	2112
	Gln	Thr	Lys	Glu	Arg	Val	Ser	Asp	Met	Asn	Leu	Ser	Gly	Thr	Ile	Thr	
			690				695					700					
	acc	cct	gct	taa													2124
	Thr	Pro	Ala														
	705																
25	<210> 61																
	<211> 707																
	<212> PRT																
	<213> Transposon Tn4451																
30	<400> 61																
	Met	Ser	Arg	Thr	Ser	Arg	Ile	Thr	Ala	Leu	Tyr	Glu	Arg	Leu	Ser	Arg	
	1				5					10					15		
35	Asp	Asp	Asp	Leu	Thr	Gly	Glu	Ser	Asn	Ser	Ile	Thr	Asn	Gln	Lys	Lys	
				20					25					30			
40	Tyr	Leu	Glu	Asp	Tyr	Ala	Arg	Arg	Asn	Gly	Phe	Glu	Asn	Ile	Arg	His	
			35					40					45				
	Phe	Thr	Asp	Asp	Gly	Phe	Ser	Gly	Val	Asn	Phe	Asn	Arg	Pro	Gly	Phe	
		50					55					60					
45	Gln	Ser	Leu	Ile	Lys	Glu	Val	Glu	Ala	Gly	Asn	Val	Glu	Thr	Leu	Ile	
	65					70					75					80	
	Val	Lys	Asp	Met	Ser	Arg	Leu	Gly	Arg	Asn	Tyr	Leu	Gln	Val	Gly	Phe	
					85					90					95		
50	Tyr	Thr	Glu	Val	Leu	Phe	Pro	Gln	Lys	Asn	Val	Arg	Phe	Leu	Ala	Ile	
				100					105					110			
55	Asn	Asn	Ser	Ile	Asp	Ser	Asn	Asn	Ala	Ser	Asp	Asn	Asp	Phe	Ala	Pro	
			115					120					125				
	Phe	Leu	Asn	Ile	Met	Asn	Glu	Trp	Tyr	Ala	Lys	Asp	Thr	Ser	Asn	Lys	
		130					135					140					
60	Ile	Lys	Ala	Ile	Phe	Asp	Ala	Arg	Met	Lys	Asp	Gly	Lys	Arg	Cys	Ser	
	145					150					155					160	
	Gly	Ser	Ile	Pro	Tyr	Gly	Tyr	Asn	Arg	Leu	Pro	Ser	Asp	Lys	Gln	Thr	
				165						170					175		
65	Leu	Val	Val	Asp	Pro	Val	Ala	Ser	Glu	Val	Val	Lys	Arg	Ile	Phe	Thr	
				180					185						190		

44

	Leu	Ala	Asn	Asp	Gly	Lys	Ser	Thr	Arg	Ala	Ile	Ala	Glu	Ile	Leu	Thr
			195					200					205			
5	Glu	Glu	Lys	Val	Leu	Thr	Pro	Ala	Ala	Tyr	Ala	Lys	Glu	Tyr	His	Pro
		210					215					220				
	Glu	Gln	Tyr	Asn	Gly	Asn	Lys	Phe	Thr	Asn	Pro	Tyr	Leu	Trp	Ala	Met
	225				230						235					240
10	Ser	Thr	Ile	Arg	Asn	Ile	Leu	Gly	Arg	Gln	Glu	Tyr	Leu	Gly	His	Thr
					245					250					255	
	Val	Leu	Arg	Lys	Ser	Val	Ser	Thr	Asn	Phe	Lys	Leu	His	Lys	Arg	Lys
15				260					265					270		
	Ser	Thr	Asp	Glu	Glu	Glu	Gln	Tyr	Val	Phe	Pro	Asn	Thr	His	Glu	Pro
			275					280					285			
20	Ile	Ile	Ser	Gln	Glu	Leu	Trp	Asp	Ser	Val	Gln	Lys	Arg	Arg	Ser	Arg
	290						295					300				
	Val	Asn	Arg	Ala	Ser	Ala	Trp	Gly	Thr	His	Ser	Asn	Arg	Leu	Ser	Gly
	305					310					315					320
25	Tyr	Leu	Tyr	Cys	Ala	Asp	Cys	Gly	Arg	Arg	Met	Thr	Leu	Gln	Thr	His
					325					330					335	
	Tyr	Ser	Lys	Lys	Asp	Gly	Ser	Val	Gln	Tyr	Ser	Tyr	Arg	Cys	Gly	Gly
30				340					345					350		
	Tyr	Ala	Ser	Arg	Val	Asn	Ser	Cys	Thr	Ser	His	Ser	Ile	Ser	Thr	Asp
			355					360					365			
35	Asn	Val	Glu	Ala	Leu	Ile	Leu	Ser	Ser	Val	Lys	Arg	Phe	Ser	Arg	Phe
	370						375					380				
	Val	Leu	Asn	Asp	Glu	Gln	Ala	Phe	Ala	Leu	Glu	Leu	Gln	Ser	Leu	Trp
	385					390					395					400
40	Asn	Glu	Lys	Gln	Glu	Glu	Lys	Pro	Lys	His	Asn	Gln	Ser	Glu	Leu	Gln
					405					410					415	
	Arg	Cys	Gln	Lys	Arg	Tyr	Asp	Glu	Leu	Ser	Thr	Leu	Val	Arg	Gly	Leu
45				420					425					430		
	Tyr	Glu	Asn	Leu	Met	Ser	Gly	Leu	Leu	Pro	Glu	Arg	Gln	Tyr	Lys	Gln
		435					440						445			
50	Leu	Met	Lys	Gln	Tyr	Asp	Asp	Glu	Gln	Ala	Glu	Leu	Glu	Thr	Lys	Met
	450						455					460				
	Glu	Thr	Met	Lys	Thr	Glu	Leu	Ala	Glu	Glu	Lys	Val	Ser	Ser	Val	Asp
	465					470					475					480
55	Ile	Lys	His	Phe	Ile	Ser	Leu	Ile	Arg	Lys	Cys	Lys	Asn	Pro	Thr	Glu
					485					490					495	
	Ile	Ser	Asp	Thr	Met	Phe	Asn	Glu	Leu	Val	Asp	Lys	Ile	Val	Val	Tyr
60				500					505					510		
	Glu	Ala	Glu	Gly	Val	Gly	Lys	Ala	Arg	Thr	Gln	Lys	Val	Asp	Ile	Tyr
			515					520					525			
65	Phe	Asn	Tyr	Val	Gly	Gln	Val	Asp	Ile	Ala	Tyr	Thr	Glu	Glu	Glu	Leu
	530						535					540				
	Ala	Glu	Ile	Glu	Thr	Gln	Lys	Glu	Gln	Glu	Glu	Gln	Gln	Arg	Leu	Ala

45

	545			550					555					560		
	Arg	Gln	Arg	Lys	Arg	Glu	Lys	Ala	Tyr	Arg	Glu	Lys	Arg	Lys	Ala	Gln
				565					570					575		
5	Lys	Ile	Ala	Glu	Asn	Gly	Gly	Glu	Ile	Val	Lys	Thr	Lys	Val	Cys	Pro
				580					585					590		
10	His	Cys	Asn	Lys	Glu	Phe	Ile	Pro	Thr	Ser	Asn	Arg	Gln	Val	Phe	Cys
			595					600					605			
	Ser	Lys	Glu	Cys	Cys	Tyr	Gln	Ala	Arg	Gln	Asp	Lys	Lys	Lys	Thr	Asp
		610					615					620				
15	Arg	Glu	Ala	Glu	Arg	Gly	Asn	His	Tyr	Tyr	Arg	Gln	Arg	Val	Cys	Ala
		625				630					635					640
	Val	Cys	Gly	Asn	Ser	Tyr	Trp	Pro	Thr	His	Ser	Gln	Gln	Lys	Phe	Cys
				645						650					655	
20	Ser	Glu	Glu	Cys	Gln	Arg	Val	Asn	His	Asn	Lys	Lys	Thr	Leu	Glu	Phe
				660					665					670		
25	Tyr	His	His	Lys	Lys	Glu	Lys	Glu	Lys	Leu	Gln	Cys	Lys	Asp	Leu	Ser
			675					680					685			
	Gln	Thr	Lys	Glu	Arg	Val	Ser	Asp	Met	Asn	Leu	Ser	Gly	Thr	Ile	Thr
		690					695					700				
30	Thr	Pro	Ala													
		705														
35	<210> 62															
	<211> 1420															
	<212> DNA															
	<213> XisA recombinase															
40	<220>															
	<221> CDS															
	<222> (1)..(1416)															
	<400> 62															
45	atg	caa	aat	cag	ggt	caa	gac	aaa	tat	caa	caa	gcc	ttt	gca	gac	tta
	Met	Gln	Asn	Gln	Gly	Gln	Asp	Lys	Tyr	Gln	Gln	Ala	Phe	Ala	Asp	Leu
	1				5					10				15		
50	gag	cca	ctt	tca	tct	acc	gac	ggc	agt	ttt	ctc	ggc	tca	agt	ctg	caa
	Glu	Pro	Leu	Ser	Ser	Thr	Asp	Gly	Ser	Phe	Leu	Gly	Ser	Ser	Leu	Gln
				20					25					30		
55	gca	cag	cag	caa	aga	gaa	cac	atg	aga	aca	aaa	gta	cta	caa	gac	cta
	Ala	Gln	Gln	Gln	Arg	Glu	His	Met	Arg	Thr	Lys	Val	Leu	Gln	Asp	Leu
			35					40					45			
60	gac	aag	gta	aat	ctg	cgt	ttg	aag	tct	gca	aag	acg	aaa	gtc	tca	gtt
	Asp	Lys	Val	Asn	Leu	Arg	Leu	Lys	Ser	Ala	Lys	Thr	Lys	Val	Ser	Val
		50					55					60				
65	cga	gaa	tct	aac	gga	agt	ctg	caa	tta	cga	gca	acg	tta	cca	att	aaa
	Arg	Glu	Ser	Asn	Gly	Ser	Leu	Gln	Leu	Arg	Ala	Thr	Leu	Pro	Ile	Lys
		65				70				75						80
70	cct	gga	gat	aag	gac	acc	aac	ggt	aca	ggc	aga	aag	caa	tac	aat	ctc
	Pro	Gly	Asp	Lys	Asp	Thr	Asn	Gly	Thr	Gly	Arg	Lys	Gln	Tyr	Asn	Leu
					85					90					95	

5	agc	ttg	aat	atc	cct	gca	aac	ttg	gat	gga	ctg	aag	acg	gct	gag	gaa	336
	Ser	Leu	Asn	Ile	Pro	Ala	Asn	Leu	Asp	Gly	Leu	Lys	Thr	Ala	Glu	Glu	100
10	gaa	gct	tat	gaa	tta	ggt	aaa	tta	atc	gct	cgg	aaa	acc	ttt	gaa	tgg	384
	Glu	Ala	Tyr	Glu	Leu	Gly	Lys	Leu	Ile	Ala	Arg	Lys	Thr	Phe	Glu	Trp	115
15	aat	gat	aaa	tat	tta	ggc	aaa	gaa	gcc	act	aaa	aaa	gat	tca	caa	aca	432
	Asn	Asp	Lys	Tyr	Leu	Gly	Lys	Glu	Ala	Thr	Lys	Lys	Asp	Ser	Gln	Thr	130
20	ata	ggt	gat	tta	cta	gaa	aaa	ttt	gca	gaa	gag	tat	ttt	aaa	acc	cat	480
	Ile	Gly	Asp	Leu	Leu	Glu	Lys	Phe	Ala	Glu	Glu	Tyr	Phe	Lys	Thr	His	145
25	aaa	cgc	acc	act	aaa	agc	gaa	cat	acc	ttt	ttt	tac	tat	ttt	tcc	cgc	528
	Lys	Arg	Thr	Thr	Lys	Ser	Glu	His	Thr	Phe	Phe	Tyr	Tyr	Phe	Ser	Arg	165
30	acc	caa	cga	tat	acc	aat	tcc	aaa	gat	tta	gca	acg	gcg	gaa	aat	ctc	576
	Thr	Gln	Arg	Tyr	Thr	Asn	Ser	Lys	Asp	Leu	Ala	Thr	Ala	Glu	Asn	Leu	180
35	atc	aat	tca	att	gag	caa	atc	gat	aaa	gaa	tgg	gcg	aga	tat	aat	gcc	624
	Ile	Asn	Ser	Ile	Glu	Gln	Ile	Asp	Lys	Glu	Trp	Ala	Arg	Tyr	Asn	Ala	195
40	gcc	aga	gcc	ata	tca	gct	ttt	tgc	ata	aca	ttc	aat	ata	gaa	att	gat	672
	Ala	Arg	Ala	Ile	Ser	Ala	Phe	Cys	Ile	Thr	Phe	Asn	Ile	Glu	Ile	Asp	210
45	ttg	tcc	cag	tat	tcc	aaa	atg	cct	gat	cgc	aat	tcg	cgc	aac	atc	ccc	720
	Leu	Ser	Gln	Tyr	Ser	Lys	Met	Pro	Asp	Arg	Asn	Ser	Arg	Asn	Ile	Pro	225
50	aca	gat	gca	gaa	ata	cta	tca	gga	att	acc	aaa	ttt	gaa	gac	tat	cta	768
	Thr	Asp	Ala	Glu	Ile	Leu	Ser	Gly	Ile	Thr	Lys	Phe	Glu	Asp	Tyr	Leu	245
55	gtt	acc	aga	gga	aat	caa	gtt	aat	gaa	gat	gta	aaa	gat	agc	tgg	caa	816
	Val	Thr	Arg	Gly	Asn	Gln	Val	Asn	Glu	Asp	Val	Lys	Asp	Ser	Trp	Gln	260
60	ctt	tgg	cgc	tgg	aca	tat	gga	atg	tta	gca	gtt	ttt	ggt	tta	cgc	ccc	864
	Leu	Trp	Arg	Trp	Thr	Tyr	Gly	Met	Leu	Ala	Val	Phe	Gly	Leu	Arg	Pro	275
65	agg	gaa	att	ttt	att	aac	cct	aat	att	gat	tgg	tgg	tta	agc	aaa	gag	912
	Arg	Glu	Ile	Phe	Ile	Asn	Pro	Asn	Ile	Asp	Trp	Trp	Leu	Ser	Lys	Glu	290
70	aat	ata	gac	ctc	aca	tgg	aaa	gta	gac	aaa	gaa	tgt	aaa	act	ggt	gaa	960
	Asn	Ile	Asp	Leu	Thr	Trp	Lys	Val	Asp	Lys	Glu	Cys	Lys	Thr	Gly	Glu	305
75	aga	caa	gca	tta	ccc	tta	cat	aaa	gaa	tgg	att	gat	gag	ttt	gat	tta	1008
	Arg	Gln	Ala	Leu	Pro	Leu	His	Lys	Glu	Trp	Ile	Asp	Glu	Phe	Asp	Leu	325
80	aga	aat	ccg	aaa	tat	tta	gaa	atg	ctg	gca	aca	gca	att	agt	aaa	aaa	1056
	Arg	Asn	Pro	Lys	Tyr	Leu	Glu	Met	Leu	Ala	Thr	Ala	Ile	Ser	Lys	Lys	340
85	gat	aaa	aca	aat	cat	gct	gaa	ata	aca	gcc	tta	act	cag	cgt	att	agt	1104
	Asp	Lys	Thr	Asn	His	Ala	Glu	Ile	Thr	Ala	Leu	Thr	Gln	Arg	Ile	Ser	345

47

	355	360	365	
5	tgg tgg ttt cgg aaa gtc gaa tta gat ttt aaa ccc tat gat tta cgt Trp Trp Phe Arg Lys Val Glu Leu Asp Phe Lys Pro Tyr Asp Leu Arg 370 375 380			1152
10	cac gcc tgg gca atc aga gcg cat att tta ggc ata cca atc aaa gcg His Ala Trp Ala Ile Arg Ala His Ile Leu Gly Ile Pro Ile Lys Ala 385 390 395 400			1200
	gcg gct gat aat ttg ggg cat agt atg cag gtt cat aca caa acc tat Ala Ala Asp Asn Leu Gly His Ser Met Gln Val His Thr Gln Thr Tyr 405 410 415			1248
15	cag cgc tgg ttc tcg cta gat atg cgg aag tta gcg att aat cag gct Gln Arg Trp Phe Ser Leu Asp Met Arg Lys Leu Ala Ile Asn Gln Ala 420 425 430			1296
20	ttg act aag agg aat gaa ttt gag gtg att agg gag gag aat gct aaa Leu Thr Lys Arg Asn Glu Phe Glu Val Ile Arg Glu Glu Asn Ala Lys 435 440 445			1344
25	ttg cag ata gaa aat gaa agg ttg agg atg gaa att gag aag tta aag Leu Gln Ile Glu Asn Glu Arg Leu Arg Met Glu Ile Glu Lys Leu Lys 450 455 460			1392
30	atg gaa ata gct tat aag aat agt tgag Met Glu Ile Ala Tyr Lys Asn Ser 465 470			1420
35	<210> 63 <211> 472 <212> PRT <213> XisA recombinase			
40	<400> 63 Met Gln Asn Gln Gly Gln Asp Lys Tyr Gln Gln Ala Phe Ala Asp Leu 1 5 10 15			
	Glu Pro Leu Ser Ser Thr Asp Gly Ser Phe Leu Gly Ser Ser Leu Gln 20 25 30			
45	Ala Gln Gln Gln Arg Glu His Met Arg Thr Lys Val Leu Gln Asp Leu 35 40 45			
	Asp Lys Val Asn Leu Arg Leu Lys Ser Ala Lys Thr Lys Val Ser Val 50 55 60			
50	Arg Glu Ser Asn Gly Ser Leu Gln Leu Arg Ala Thr Leu Pro Ile Lys 65 70 75 80			
55	Pro Gly Asp Lys Asp Thr Asn Gly Thr Gly Arg Lys Gln Tyr Asn Leu 85 90 95			
	Ser Leu Asn Ile Pro Ala Asn Leu Asp Gly Leu Lys Thr Ala Glu Glu 100 105 110			
60	Glu Ala Tyr Glu Leu Gly Lys Leu Ile Ala Arg Lys Thr Phe Glu Trp 115 120 125			
	Asn Asp Lys Tyr Leu Gly Lys Glu Ala Thr Lys Lys Asp Ser Gln Thr 130 135 140			
65	Ile Gly Asp Leu Leu Glu Lys Phe Ala Glu Glu Tyr Phe Lys Thr His 145 150 155 160			

48

	Lys	Arg	Thr	Thr	Lys	Ser	Glu	His	Thr	Phe	Phe	Tyr	Tyr	Phe	Ser	Arg
					165					170					175	
5	Thr	Gln	Arg	Tyr	Thr	Asn	Ser	Lys	Asp	Leu	Ala	Thr	Ala	Glu	Asn	Leu
				180					185					190		
	Ile	Asn	Ser	Ile	Glu	Gln	Ile	Asp	Lys	Glu	Trp	Ala	Arg	Tyr	Asn	Ala
			195					200					205			
10	Ala	Arg	Ala	Ile	Ser	Ala	Phe	Cys	Ile	Thr	Phe	Asn	Ile	Glu	Ile	Asp
		210					215					220				
	Leu	Ser	Gln	Tyr	Ser	Lys	Met	Pro	Asp	Arg	Asn	Ser	Arg	Asn	Ile	Pro
15		225				230					235					240
	Thr	Asp	Ala	Glu	Ile	Leu	Ser	Gly	Ile	Thr	Lys	Phe	Glu	Asp	Tyr	Leu
				245						250					255	
20	Val	Thr	Arg	Gly	Asn	Gln	Val	Asn	Glu	Asp	Val	Lys	Asp	Ser	Trp	Gln
				260					265					270		
	Leu	Trp	Arg	Trp	Thr	Tyr	Gly	Met	Leu	Ala	Val	Phe	Gly	Leu	Arg	Pro
			275					280					285			
25	Arg	Glu	Ile	Phe	Ile	Asn	Pro	Asn	Ile	Asp	Trp	Trp	Leu	Ser	Lys	Glu
		290					295					300				
	Asn	Ile	Asp	Leu	Thr	Trp	Lys	Val	Asp	Lys	Glu	Cys	Lys	Thr	Gly	Glu
30		305				310					315					320
	Arg	Gln	Ala	Leu	Pro	Leu	His	Lys	Glu	Trp	Ile	Asp	Glu	Phe	Asp	Leu
					325					330					335	
35	Arg	Asn	Pro	Lys	Tyr	Leu	Glu	Met	Leu	Ala	Thr	Ala	Ile	Ser	Lys	Lys
				340					345					350		
	Asp	Lys	Thr	Asn	His	Ala	Glu	Ile	Thr	Ala	Leu	Thr	Gln	Arg	Ile	Ser
			355					360					365			
40	Trp	Trp	Phe	Arg	Lys	Val	Glu	Leu	Asp	Phe	Lys	Pro	Tyr	Asp	Leu	Arg
		370					375					380				
	His	Ala	Trp	Ala	Ile	Arg	Ala	His	Ile	Leu	Gly	Ile	Pro	Ile	Lys	Ala
45		385				390					395					400
	Ala	Ala	Asp	Asn	Leu	Gly	His	Ser	Met	Gln	Val	His	Thr	Gln	Thr	Tyr
					405					410					415	
50	Gln	Arg	Trp	Phe	Ser	Leu	Asp	Met	Arg	Lys	Leu	Ala	Ile	Asn	Gln	Ala
			420						425					430		
	Leu	Thr	Lys	Arg	Asn	Glu	Phe	Glu	Val	Ile	Arg	Glu	Glu	Asn	Ala	Lys
			435					440					445			
55	Leu	Gln	Ile	Glu	Asn	Glu	Arg	Leu	Arg	Met	Glu	Ile	Glu	Lys	Leu	Lys
		450					455					460				
	Met	Glu	Ile	Ala	Tyr	Lys	Asn	Ser								
60		465				470										
	<210>	64														
	<211>	1008														
65	<212>	DNA														
	<213>	Artificial Sequence														

<220>

<221> CDS

<222> (1)..(1005)

5

<220>

<223> Description of Artificial Sequence: vector
pBS-SSV3

<400> 64

10	atg acg aaa gat aag acg cgt tat aaa tac ggg gat tat att tta cgc	48
	Met Thr Lys Asp Lys Thr Arg Tyr Lys Tyr Gly Asp Tyr Ile Leu Arg	
	1 5 10 15	
15	gag agg aaa ggg cgg tat tat gtt tac aag cta gag tat gaa aac ggt	96
	Glu Arg Lys Gly Arg Tyr Tyr Val Tyr Lys Leu Glu Tyr Glu Asn Gly	
	20 25 30	
20	gag gta aaa gag cgt tac gtg ggt cct tta gct gac gtc gtt gaa tca	144
	Glu Val Lys Glu Arg Tyr Val Gly Pro Leu Ala Asp Val Val Glu Ser	
	35 40 45	
25	tat cta aaa atg aaa tta ggg gtc gta ggg gat act ccc cta caa gcg	192
	Tyr Leu Lys Met Lys Leu Gly Val Val Gly Asp Thr Pro Leu Gln Ala	
	50 55 60	
30	gat ccc ccc ggt ttc gag ccc ggg aca agc gga agc ggt ggt gga aaa	240
	Asp Pro Pro Gly Phe Glu Pro Gly Thr Ser Gly Ser Gly Gly Gly Lys	
	65 70 75 80	
35	gag gga act gaa cga cgt aaa ata gcg ttg gtt gcc aat ttg cgc caa	288
	Glu Gly Thr Glu Arg Arg Lys Ile Ala Leu Val Ala Asn Leu Arg Gln	
	85 90 95	
40	tac gcg acg gac ggc aac ata aag gcg ttc tac aac tat ctc atg aac	336
	Tyr Ala Thr Asp Gly Asn Ile Lys Ala Phe Tyr Asn Tyr Leu Met Asn	
	100 105 110	
45	gaa agg ggg ata agc gaa aaa act gca aag gac tac atc aat gct ata	384
	Glu Arg Gly Ile Ser Glu Lys Thr Ala Lys Asp Tyr Ile Asn Ala Ile	
	115 120 125	
50	tca aag ccg tat aaa gag acg aga gac gca cag aag gct tac cga ctc	432
	Ser Lys Pro Tyr Lys Glu Thr Arg Asp Ala Gln Lys Ala Tyr Arg Leu	
	130 135 140	
55	ttt gca cgt ttc tta gcg tca cgc aat atc ata cat gat gaa ttt gcg	480
	Phe Ala Arg Phe Leu Ala Ser Arg Asn Ile Ile His Asp Glu Phe Ala	
	145 150 155 160	
60	gat aaa ata ttg aaa gcg gta aag gtg aag aag gcg aac gct gat atc	528
	Asp Lys Ile Leu Lys Ala Val Lys Val Lys Lys Ala Asn Ala Asp Ile	
	165 170 175	
65	tac att cca acg ttg gaa gag ata aaa agg acg tta caa tta gca aaa	576
	Tyr Ile Pro Thr Leu Glu Glu Ile Lys Arg Thr Leu Gln Leu Ala Lys	
	180 185 190	
70	gac tat agc gaa aac gtc tac ttc atc tac cgt atc gct ctc gag tcg	624
	Asp Tyr Ser Glu Asn Val Tyr Phe Ile Tyr Arg Ile Ala Leu Glu Ser	
	195 200 205	
75	ggc gtt agg ctg agc gaa ata ctg aaa gtg ctg aag gaa ccc gaa agg	672
	Gly Val Arg Leu Ser Glu Ile Leu Lys Val Leu Lys Glu Pro Glu Arg	
	210 215 220	
80	gac att tgc ggt aac gac gtc tgt tat tat ccg ctt agt tgg act agg	720
	Asp Ile Cys Gly Asn Asp Val Cys Tyr Tyr Pro Leu Ser Trp Thr Arg	

	50																
	225				230				235				240				
5	gga	tat	aag	ggc	gtc	ttc	tat	gta	ttc	cac	ata	acg	cct	ctg	aag	aga	768
	Gly	Tyr	Lys	Gly	Val	Phe	Tyr	Val	Phe	His	Ile	Thr	Pro	Leu	Lys	Arg	
					245					250					255		
10	gta	gag	gtg	acg	aag	tgg	gca	ata	gcg	gac	ttt	gaa	cga	cgt	cat	aag	816
	Val	Glu	Val	Thr	Lys	Trp	Ala	Ile	Ala	Asp	Phe	Glu	Arg	Arg	His	Lys	
				260					265					270			
15	gac	gct	ata	gcg	ata	aag	tac	ttc	cgc	aaa	ttc	gta	gcg	tct	aag	atg	864
	Asp	Ala	Ile	Ala	Ile	Lys	Tyr	Phe	Arg	Lys	Phe	Val	Ala	Ser	Lys	Met	
			275					280					285				
20	gct	gag	cta	agc	gta	ccg	tta	gat	att	atc	gat	ttt	att	caa	ggg	cgt	912
	Ala	Glu	Leu	Ser	Val	Pro	Leu	Asp	Ile	Ile	Asp	Phe	Ile	Gln	Gly	Arg	
		290					295					300					
25	aaa	ccg	aca	cgc	gtt	tta	acg	caa	cat	tac	gta	tcg	ctc	ttc	ggc	ata	960
	Lys	Pro	Thr	Arg	Val	Leu	Thr	Gln	His	Tyr	Val	Ser	Leu	Phe	Gly	Ile	
	305					310					315					320	
30	gcg	aaa	gag	caa	tat	aaa	aag	tat	gcg	gaa	tgg	cta	aaa	ggg	gtc	tga	1008
	Ala	Lys	Glu	Gln	Tyr	Lys	Lys	Tyr	Ala	Glu	Trp	Leu	Lys	Gly	Val		
					325					330					335		
35	<210> 65																
	<211> 335																
	<212> PRT																
	<213> Artificial Sequence																
	<223> Description of Artificial Sequence: vector																
	pBS-SSV3																
40	<400> 65																
	Met	Thr	Lys	Asp	Lys	Thr	Arg	Tyr	Lys	Tyr	Gly	Asp	Tyr	Ile	Leu	Arg	
	1				5					10					15		
45	Glu	Arg	Lys	Gly	Arg	Tyr	Tyr	Val	Tyr	Lys	Leu	Glu	Tyr	Glu	Asn	Gly	
				20					25					30			
50	Glu	Val	Lys	Glu	Arg	Tyr	Val	Gly	Pro	Leu	Ala	Asp	Val	Val	Glu	Ser	
			35					40					45				
55	Tyr	Leu	Lys	Met	Lys	Leu	Gly	Val	Val	Gly	Asp	Thr	Pro	Leu	Gln	Ala	
		50					55					60					
60	Asp	Pro	Pro	Gly	Phe	Glu	Pro	Gly	Thr	Ser	Gly	Ser	Gly	Gly	Gly	Lys	
	65					70					75					80	
65	Glu	Gly	Thr	Glu	Arg	Arg	Lys	Ile	Ala	Leu	Val	Ala	Asn	Leu	Arg	Gln	
					85					90					95		
70	Tyr	Ala	Thr	Asp	Gly	Asn	Ile	Lys	Ala	Phe	Tyr	Asn	Tyr	Leu	Met	Asn	
				100					105						110		
75	Glu	Arg	Gly	Ile	Ser	Glu	Lys	Thr	Ala	Lys	Asp	Tyr	Ile	Asn	Ala	Ile	
			115					120					125				
80	Ser	Lys	Pro	Tyr	Lys	Glu	Thr	Arg	Asp	Ala	Gln	Lys	Ala	Tyr	Arg	Leu	
		130					135					140					
85	Phe	Ala	Arg	Phe	Leu	Ala	Ser	Arg	Asn	Ile	Ile	His	Asp	Glu	Phe	Ala	
	145					150					155					160	
90	Asp	Lys	Ile	Leu	Lys	Ala	Val	Lys	Val	Lys	Lys	Ala	Asn	Ala	Asp	Ile	
					165					170					175		

Tyr Ile Pro Thr Leu Glu Glu Ile Lys Arg Thr Leu Gln Leu Ala Lys
 180 185 190
 5 Asp Tyr Ser Glu Asn Val Tyr Phe Ile Tyr Arg Ile Ala Leu Glu Ser
 195 200 205
 Gly Val Arg Leu Ser Glu Ile Leu Lys Val Leu Lys Glu Pro Glu Arg
 210 215 220
 10 Asp Ile Cys Gly Asn Asp Val Cys Tyr Tyr Pro Leu Ser Trp Thr Arg
 225 230 235 240
 Gly Tyr Lys Gly Val Phe Tyr Val Phe His Ile Thr Pro Leu Lys Arg
 245 250 255
 15 Val Glu Val Thr Lys Trp Ala Ile Ala Asp Phe Glu Arg Arg His Lys
 260 265 270
 20 Asp Ala Ile Ala Ile Lys Tyr Phe Arg Lys Phe Val Ala Ser Lys Met
 275 280 285
 Ala Glu Leu Ser Val Pro Leu Asp Ile Ile Asp Phe Ile Gln Gly Arg
 290 295 300
 25 Lys Pro Thr Arg Val Leu Thr Gln His Tyr Val Ser Leu Phe Gly Ile
 305 310 315 320
 30 Ala Lys Glu Gln Tyr Lys Lys Tyr Ala Glu Trp Leu Lys Gly Val
 325 330 335
 35 <210> 66
 <211> 1441
 <212> DNA
 <213> Artificial Sequence
 40 <220>
 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein NLS-XisA
 <220>
 <221> CDS
 45 <222> (1)..(1437)
 <400> 66
 atg ccc aag aag aag agg aag gtg caa aat cag ggt caa gac aaa tat 48
 Met Pro Lys Lys Lys Arg Lys Val Gln Asn Gln Gly Gln Asp Lys Tyr
 50 1 5 10 15
 caa caa gcc ttt gca gac tta gag cca ctt tca tct acc gac ggc agt 96
 Gln Gln Ala Phe Ala Asp Leu Glu Pro Leu Ser Ser Thr Asp Gly Ser
 20 25 30
 55 ttt ctc ggc tca agt ctg caa gca cag cag caa aga gaa cac atg aga 144
 Phe Leu Gly Ser Ser Leu Gln Ala Gln Gln Gln Arg Glu His Met Arg
 35 40 45
 60 aca aaa gta cta caa gac cta gac aag gta aat ctg cgt ttg aag tct 192
 Thr Lys Val Leu Gln Asp Leu Asp Lys Val Asn Leu Arg Leu Lys Ser
 50 55 60
 65 gca aag acg aaa gtc tca gtt cga gaa tct aac gga agt ctg caa tta 240
 Ala Lys Thr Lys Val Ser Val Arg Glu Ser Asn Gly Ser Leu Gln Leu
 65 70 75 80

52

	cga	gca	acg	tta	cca	att	aaa	cct	gga	gat	aag	gac	acc	aac	ggg	aca	288
	Arg	Ala	Thr	Leu	Pro	Ile	Lys	Pro	Gly	Asp	Lys	Asp	Thr	Asn	Gly	Thr	
					85					90					95		
5	ggc	aga	aag	caa	tac	aat	ctc	agc	ttg	aat	atc	cct	gca	aac	ttg	gat	336
	Gly	Arg	Lys	Gln	Tyr	Asn	Leu	Ser	Leu	Asn	Ile	Pro	Ala	Asn	Leu	Asp	
				100					105					110			
10	gga	ctg	aag	acg	gct	gag	gaa	gaa	gct	tat	gaa	tta	ggg	aaa	tta	atc	384
	Gly	Leu	Lys	Thr	Ala	Glu	Glu	Glu	Ala	Tyr	Glu	Leu	Gly	Lys	Leu	Ile	
			115					120					125				
15	gct	cgg	aaa	acc	ttt	gaa	tgg	aat	gat	aaa	tat	tta	ggc	aaa	gaa	gcc	432
	Ala	Arg	Lys	Thr	Phe	Glu	Trp	Asn	Asp	Lys	Tyr	Leu	Gly	Lys	Glu	Ala	
		130					135					140					
20	act	aaa	aaa	gat	tca	caa	aca	ata	ggg	gat	tta	cta	gaa	aaa	ttt	gca	480
	Thr	Lys	Lys	Asp	Ser	Gln	Thr	Ile	Gly	Asp	Leu	Leu	Glu	Lys	Phe	Ala	
	145					150					155					160	
	gaa	gag	tat	ttt	aaa	acc	cat	aaa	cgc	acc	act	aaa	agc	gaa	cat	acc	528
	Glu	Glu	Tyr	Phe	Lys	Thr	His	Lys	Arg	Thr	Thr	Lys	Ser	Glu	His	Thr	
					165					170					175		
25	ttt	ttt	tac	tat	ttt	tcc	cgc	acc	caa	cga	tat	acc	aat	tcc	aaa	gat	576
	Phe	Phe	Tyr	Tyr	Phe	Ser	Arg	Thr	Gln	Arg	Tyr	Thr	Asn	Ser	Lys	Asp	
				180					185					190			
30	tta	gca	acg	gcg	gaa	aat	ctc	atc	aat	tca	att	gag	caa	atc	gat	aaa	624
	Leu	Ala	Thr	Ala	Glu	Asn	Leu	Ile	Asn	Ser	Ile	Glu	Gln	Ile	Asp	Lys	
				195				200					205				
35	gaa	tgg	gcg	aga	tat	aat	gcc	gcc	aga	gcc	ata	tca	gct	ttt	tgc	ata	672
	Glu	Trp	Ala	Arg	Tyr	Asn	Ala	Ala	Arg	Ala	Ile	Ser	Ala	Phe	Cys	Ile	
		210					215					220					
40	aca	ttc	aat	ata	gaa	att	gat	ttg	tcc	cag	tat	tcc	aaa	atg	cct	gat	720
	Thr	Phe	Asn	Ile	Glu	Ile	Asp	Leu	Ser	Gln	Tyr	Ser	Lys	Met	Pro	Asp	
	225					230					235					240	
	cgc	aat	tcg	cgc	aac	atc	ccc	aca	gat	gca	gaa	ata	cta	tca	gga	att	768
	Arg	Asn	Ser	Arg	Asn	Ile	Pro	Thr	Asp	Ala	Glu	Ile	Leu	Ser	Gly	Ile	
					245					250					255		
45	acc	aaa	ttt	gaa	gac	tat	cta	gtt	acc	aga	gga	aat	caa	gtt	aat	gaa	816
	Thr	Lys	Phe	Glu	Asp	Tyr	Leu	Val	Thr	Arg	Gly	Asn	Gln	Val	Asn	Glu	
				260				265					270				
50	gat	gta	aaa	gat	agc	tgg	caa	ctt	tgg	cgc	tgg	aca	tat	gga	atg	tta	864
	Asp	Val	Lys	Asp	Ser	Trp	Gln	Leu	Trp	Arg	Trp	Thr	Tyr	Gly	Met	Leu	
			275				280						285				
55	gca	gtt	ttt	ggg	tta	cgc	ccc	agg	gaa	att	ttt	att	aac	cct	aat	att	912
	Ala	Val	Phe	Gly	Leu	Arg	Pro	Arg	Glu	Ile	Phe	Ile	Asn	Pro	Asn	Ile	
		290					295					300					
60	gat	tgg	tgg	tta	agc	aaa	gag	aat	ata	gac	ctc	aca	tgg	aaa	gta	gac	960
	Asp	Trp	Trp	Leu	Ser	Lys	Glu	Asn	Ile	Asp	Leu	Thr	Trp	Lys	Val	Asp	
	305					310					315				320		
	aaa	gaa	tgt	aaa	act	ggg	gaa	aga	caa	gca	tta	ccc	tta	cat	aaa	gaa	1008
	Lys	Glu	Cys	Lys	Thr	Gly	Glu	Arg	Gln	Ala	Leu	Pro	Leu	His	Lys	Glu	
					325					330					335		
65	tgg	att	gat	gag	ttt	gat	tta	aga	aat	cgg	aaa	tat	tta	gaa	atg	ctg	1056
	Trp	Ile	Asp	Glu	Phe	Asp	Leu	Arg	Asn	Pro	Lys	Tyr	Leu	Glu	Met	Leu	
				340					345					350			

5	gca aca gca att agt aaa aaa gat aaa aca aat cat gct gaa ata aca	1104
	Ala Thr Ala Ile Ser Lys Lys Asp Lys Thr Asn His Ala Glu Ile Thr	
	355 360 365	
10	gcc tta act cag cgt att agt tgg tgg ttt cgg aaa gtc gaa tta gat	1152
	Ala Leu Thr Gln Arg Ile Ser Trp Trp Phe Arg Lys Val Glu Leu Asp	
	370 375 380	
15	ttt aaa ccc tat gat tta cgt cac gcc tgg gca atc aga gcg cat att	1200
	Phe Lys Pro Tyr Asp Leu Arg His Ala Trp Ala Ile Arg Ala His Ile	
	385 390 395 400	
20	tta ggc ata cca atc aaa gcg gcg gct gat aat ttg ggg cat agt atg	1248
	Leu Gly Ile Pro Ile Lys Ala Ala Ala Asp Asn Leu Gly His Ser Met	
	405 410 415	
25	cag gtt cat aca caa acc tat cag cgc tgg ttc tcg cta gat atg cgg	1296
	Gln Val His Thr Gln Thr Tyr Gln Arg Trp Phe Ser Leu Asp Met Arg	
	420 425 430	
30	aag tta gcg att aat cag gct ttg act aag agg aat gaa ttt gag gtg	1344
	Lys Leu Ala Ile Asn Gln Ala Leu Thr Lys Arg Asn Glu Phe Glu Val	
	435 440 445	
35	att agg gag gag aat gct aaa ttg cag ata gaa aat gaa agg ttg agg	1392
	Ile Arg Glu Glu Asn Ala Lys Leu Gln Ile Glu Asn Glu Arg Leu Arg	
	450 455 460	
40	atg gaa att gag aag tta aag atg gaa ata gct tat aag aat agt tgag	1441
	Met Glu Ile Glu Lys Leu Lys Met Glu Ile Ala Tyr Lys Asn Ser	
	465 470 475	
45	<210> 67	
	<211> 479	
50	<212> PRT	
	<213> Artificial Sequence	
55	<223> Description of Artificial Sequence: DNA sequence	
	coding for fusion protein NLS-XisA	
60	<400> 67	
	Met Pro Lys Lys Lys Arg Lys Val Gln Asn Gln Gly Gln Asp Lys Tyr	
	1 5 10 15	
65	Gln Gln Ala Phe Ala Asp Leu Glu Pro Leu Ser Ser Thr Asp Gly Ser	
	20 25 30	
70	Phe Leu Gly Ser Ser Leu Gln Ala Gln Gln Gln Arg Glu His Met Arg	
	35 40 45	
75	Thr Lys Val Leu Gln Asp Leu Asp Lys Val Asn Leu Arg Leu Lys Ser	
	50 55 60	
80	Ala Lys Thr Lys Val Ser Val Arg Glu Ser Asn Gly Ser Leu Gln Leu	
	65 70 75 80	
85	Arg Ala Thr Leu Pro Ile Lys Pro Gly Asp Lys Asp Thr Asn Gly Thr	
	85 90 95	
90	Gly Arg Lys Gln Tyr Asn Leu Ser Leu Asn Ile Pro Ala Asn Leu Asp	
	100 105 110	
95	Gly Leu Lys Thr Ala Glu Glu Glu Ala Tyr Glu Leu Gly Lys Leu Ile	
	115 120 125	
100	Ala Arg Lys Thr Phe Glu Trp Asn Asp Lys Tyr Leu Gly Lys Glu Ala	

<210> 68
 <211> 1029
 <212> DNA
 <213> Artificial Sequence

5

<220>
 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein NLS-Ssv

10

<220>
 <221> CDS
 <222> (1)..(1026)

<400> 68
 15 atg ccc aag aag aag agg aag gtg acg aaa gat aag acg cgt tat aaa 48
 Met Pro Lys Lys Lys Arg Lys Val Thr Lys Asp Lys Thr Arg Tyr Lys
 1 5 10 15
 20 tac ggg gat tat att tta cgc gag agg aaa ggg cgg tat tat gtt tac 96
 Tyr Gly Asp Tyr Ile Leu Arg Glu Arg Lys Gly Arg Tyr Tyr Val Tyr
 20 25 30
 25 aag cta gag tat gaa aac ggt gag gta aaa gag cgt tac gtg ggt cct 144
 Lys Leu Glu Tyr Glu Asn Gly Glu Val Lys Glu Arg Tyr Val Gly Pro
 35 40 45
 30 tta gct gac gtc gtt gaa tca tat cta aaa atg aaa tta ggg gtc gta 192
 Leu Ala Asp Val Val Glu Ser Tyr Leu Lys Met Lys Leu Gly Val Val
 50 55 60
 35 ggg gat act ccc cta caa gcg gat ccc ccc ggt ttc gag ccc ggg aca 240
 Gly Asp Thr Pro Leu Gln Ala Asp Pro Pro Gly Phe Glu Pro Gly Thr
 65 70 75 80
 40 agc gga agc ggt ggt gga aaa gag gga act gaa cga cgt aaa ata gcg 288
 Ser Gly Ser Gly Gly Gly Lys Glu Gly Thr Glu Arg Arg Lys Ile Ala
 85 90 95
 45 ttg gtt gcc aat ttg cgc caa tac gcg acg gac ggc aac ata aag gcg 336
 Leu Val Ala Asn Leu Arg Gln Tyr Ala Thr Asp Gly Asn Ile Lys Ala
 100 105 110
 50 ttc tac aac tat ctc atg aac gaa agg ggg ata agc gaa aaa act gca 384
 Phe Tyr Asn Tyr Leu Met Asn Glu Arg Gly Ile Ser Glu Lys Thr Ala
 115 120 125
 55 aag gac tac atc aat gct ata tca aag ccg tat aaa gag acg aga gac 432
 Lys Asp Tyr Ile Asn Ala Ile Ser Lys Pro Tyr Lys Glu Thr Arg Asp
 130 135 140
 60 gca cag aag gct tac cga ctc ttt gca cgt ttc tta gcg tca cgc aat 480
 Ala Gln Lys Ala Tyr Arg Leu Phe Ala Arg Phe Leu Ala Ser Arg Asn
 145 150 155 160
 65 atc ata cat gat gaa ttt gcg gat aaa ata ttg aaa gcg gta aag gtg 528
 Ile Ile His Asp Glu Phe Ala Asp Lys Ile Leu Lys Ala Val Lys Val
 165 170 175
 70 aag aag gcg aac gct gat atc tac att cca acg ttg gaa gag ata aaa 576
 Lys Lys Ala Asn Ala Asp Ile Tyr Ile Pro Thr Leu Glu Glu Ile Lys
 180 185 190
 75 agg acg tta caa tta gca aaa gac tat agc gaa aac gtc tac ttc atc 624
 Arg Thr Leu Gln Leu Ala Lys Asp Tyr Ser Glu Asn Val Tyr Phe Ile
 195 200 205
 80 tac cgt atc gct ctc gag tcg ggc gtt agg ctg agc gaa ata ctg aaa 672

56

	Tyr	Arg	Ile	Ala	Leu	Glu	Ser	Gly	Val	Arg	Leu	Ser	Glu	Ile	Leu	Lys	
	210						215					220					
5	gtg	ctg	aag	gaa	ccc	gaa	agg	gac	att	tgc	ggg	aac	gac	gtc	tgt	tat	720
	Val	Leu	Lys	Glu	Pro	Glu	Arg	Asp	Ile	Cys	Gly	Asn	Asp	Val	Cys	Tyr	
	225					230					235					240	
10	tat	ccg	ctt	agt	tgg	act	agg	gga	tat	aag	ggc	gtc	ttc	tat	gta	ttc	768
	Tyr	Pro	Leu	Ser	Trp	Thr	Arg	Gly	Tyr	Lys	Gly	Val	Phe	Tyr	Val	Phe	
					245					250					255		
15	cac	ata	acg	cct	ctg	aag	aga	gta	gag	gtg	acg	aag	tgg	gca	ata	gcg	816
	His	Ile	Thr	Pro	Leu	Lys	Arg	Val	Glu	Val	Thr	Lys	Trp	Ala	Ile	Ala	
				260					265					270			
	gac	ttt	gaa	cga	cgt	cat	aag	gac	gct	ata	gcg	ata	aag	tac	ttc	cgc	864
	Asp	Phe	Glu	Arg	Arg	His	Lys	Asp	Ala	Ile	Ala	Ile	Lys	Tyr	Phe	Arg	
			275					280					285				
20	aaa	ttc	gta	gcg	tct	aag	atg	gct	gag	cta	agc	gta	ccg	tta	gat	att	912
	Lys	Phe	Val	Ala	Ser	Lys	Met	Ala	Glu	Leu	Ser	Val	Pro	Leu	Asp	Ile	
		290					295					300					
25	atc	gat	ttt	att	caa	ggg	cgt	aaa	ccg	aca	cgc	gtt	tta	acg	caa	cat	960
	Ile	Asp	Phe	Ile	Gln	Gly	Arg	Lys	Pro	Thr	Arg	Val	Leu	Thr	Gln	His	
	305					310					315					320	
30	tac	gta	tcg	ctc	ttc	ggc	ata	gcg	aaa	gag	caa	tat	aaa	aag	tat	gcg	1008
	Tyr	Val	Ser	Leu	Phe	Gly	Ile	Ala	Lys	Glu	Gln	Tyr	Lys	Lys	Tyr	Ala	
					325					330					335		
35	gaa	tgg	cta	aaa	ggg	gtc	tga										1029
	Glu	Trp	Leu	Lys	Gly	Val											
				340													
40	<210> 69 <211> 342 <212> PRT <213> Artificial Sequence <223> Description of Artificial Sequence: DNA sequence coding for fusion protein NLS-Ssv																
45	<400> 69 Met Pro Lys Lys Lys Arg Lys Val Thr Lys Asp Lys Thr Arg Tyr Lys 1 5 10 15																
50	Tyr	Gly	Asp	Tyr	Ile	Leu	Arg	Glu	Arg	Lys	Gly	Arg	Tyr	Tyr	Val	Tyr	
				20					25					30			
	Lys	Leu	Glu	Tyr	Glu	Asn	Gly	Glu	Val	Lys	Glu	Arg	Tyr	Val	Gly	Pro	
			35					40					45				
55	Leu	Ala	Asp	Val	Val	Glu	Ser	Tyr	Leu	Lys	Met	Lys	Leu	Gly	Val	Val	
		50					55					60					
	Gly	Asp	Thr	Pro	Leu	Gln	Ala	Asp	Pro	Pro	Gly	Phe	Glu	Pro	Gly	Thr	
	65					70					75					80	
60	Ser	Gly	Ser	Gly	Gly	Gly	Lys	Glu	Gly	Thr	Glu	Arg	Arg	Lys	Ile	Ala	
					85				90						95		
65	Leu	Val	Ala	Asn	Leu	Arg	Gln	Tyr	Ala	Thr	Asp	Gly	Asn	Ile	Lys	Ala	
				100					105					110			
	Phe	Tyr	Asn	Tyr	Leu	Met	Asn	Glu	Arg	Gly	Ile	Ser	Glu	Lys	Thr	Ala	
			115					120					125				

Lys Asp Tyr Ile Asn Ala Ile Ser Lys Pro Tyr Lys Glu Thr Arg Asp
 130 135 140
 5 Ala Gln Lys Ala Tyr Arg Leu Phe Ala Arg Phe Leu Ala Ser Arg Asn
 145 150 155 160
 Ile Ile His Asp Glu Phe Ala Asp Lys Ile Leu Lys Ala Val Lys Val
 165 170 175
 10 Lys Lys Ala Asn Ala Asp Ile Tyr Ile Pro Thr Leu Glu Glu Ile Lys
 180 185 190
 15 Arg Thr Leu Gln Leu Ala Lys Asp Tyr Ser Glu Asn Val Tyr Phe Ile
 195 200 205
 Tyr Arg Ile Ala Leu Glu Ser Gly Val Arg Leu Ser Glu Ile Leu Lys
 210 215 220
 20 Val Leu Lys Glu Pro Glu Arg Asp Ile Cys Gly Asn Asp Val Cys Tyr
 225 230 235 240
 Tyr Pro Leu Ser Trp Thr Arg Gly Tyr Lys Gly Val Phe Tyr Val Phe
 245 250 255
 25 His Ile Thr Pro Leu Lys Arg Val Glu Val Thr Lys Trp Ala Ile Ala
 260 265 270
 30 Asp Phe Glu Arg Arg His Lys Asp Ala Ile Ala Ile Lys Tyr Phe Arg
 275 280 285
 Lys Phe Val Ala Ser Lys Met Ala Glu Leu Ser Val Pro Leu Asp Ile
 290 295 300
 35 Ile Asp Phe Ile Gln Gly Arg Lys Pro Thr Arg Val Leu Thr Gln His
 305 310 315 320
 Tyr Val Ser Leu Phe Gly Ile Ala Lys Glu Gln Tyr Lys Lys Tyr Ala
 325 330 335
 40 Glu Trp Leu Lys Gly Val
 340
 45 <210> 70
 <211> 3908
 <212> DNA
 <213> Artificial Sequence
 50 <220>
 <223> Description of Artificial Sequence: vector
 pBS-SSV3
 55 <400> 70
 cacctaaatt gtaagcggtta atattttggtt aaaatttcgog tttaaattttt gttaaatcag 60
 ctcatattttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa agaatagac 120
 cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga 180
 ctccaacgtc aaaggcgcaa aaaccgtcta tcaggggcgt ggcccactac gtgaaccatc 240
 60 accctaataca agtttttttg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg 300
 gagccccga tttagagctt gacggggaaa gccggcgaaac gtggcgagaa aggaagggaa 360
 gaaagcgaaa ggagcgggct ctagggcgct ggcaagtgt gcggtcacgc tgcgcgtaac 420
 caccacaccc gccgcgctta atgcgcgct acaggcgcg tccattcgc cattcaggct 480
 ggcgaactgt tgggaagggc gatcgggtgc ggcctcttcg ctattacgcc agctggcgaa 540
 65 agggggatgt gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg 600
 ttgtaaaacg acggccagt aattgtaata cgactcacta tagggcgaat tggagctcca 660
 ccgcggtggc ggcggccga tatgacgaaa gataagacgc gttataaata cggggattat 720

58

	attttacgcg	agaggaaagg	gcggtattat	gtttacaagc	tagagtatga	aaacgggtgag	780
	gtaaaagagc	gttacgtggg	tccttttagct	gacgtcgttg	aatcatatct	aaaaatgaaa	840
	ttaggggtcg	taggggatac	tccctacaa	gcggatcccc	ccggtttcga	gcccgggaca	900
5	agcggaagcg	gtggtggaaa	agagggaact	gaacgacgta	aaatagcgtt	ggttgccaat	960
	ttgcgccaat	acgcgacgga	cggcaacata	aaggcgttct	acaactatct	catgaacgaa	1020
	agggggataa	gcgaaaaaac	tgcaaaaggac	tacatcaatg	ctatatcaaa	gccgtataaa	1080
	gagacgagag	acgcacagaa	ggcttaccga	ctctttgcac	gtttcttagc	gtcacgcaat	1140
	atcatatctg	atgaatttgc	ggataaaaata	ttgaaagcgg	taaagggtgaa	gaaggcgaa	1200
10	gctgatattct	acattccaac	gttggaagag	ataaaaaggga	cgttacaatt	agcaaaagac	1260
	tatagcgaaa	acgtctactt	catctaccgt	atcgctctcg	agtcgggogt	taggotgagc	1320
	gaaatactga	aagtgtctgaa	ggaacccgaa	agggacattt	gcggtaacga	cgtctgttat	1380
	tatccgctta	gttggactag	gggatataag	ggcgtcttct	atgtattcca	cataacgcct	1440
	ctgaagagag	tagagggtgac	gaagtgggca	atagcggact	ttgaacgacg	tcataaggac	1500
15	gctatagcga	taaagtactt	ccgcaaatc	gtgacgtcta	agatggctga	gctaagcgta	1560
	ccgttagata	ttatcgattt	tattcaaggg	cgtaaaccca	cacgcgtttt	aacgcaacat	1620
	tacgtatcgc	tcttcggcat	agcgaaagag	caatataaaa	agtatgcgga	atggctaaaa	1680
	ggggtctgac	tcgagggggg	gcccggtacc	cagcttttgt	tcccttttagt	gagggttaat	1740
	ttcgagcttg	gcgtaatcat	ggctcatagct	gtttcctgtg	tgaattgtt	atccgctcac	1800
20	aattccacac	aacatacagag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaattgagt	1860
	gagctaaactc	acattaattg	cgttgcgctc	actgccgct	ttccagtcgg	gaaacctgtc	1920
	gtgccagctg	cattaatgaa	tcggccaacg	cgcggggaga	ggcggtttgc	gtattgggcg	1980
	ctcttcgct	tcctcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggt	2040
	atcagctcac	tcaaaggcgg	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	2100
25	gaacatgtga	gcaaaaggcc	agcaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	2160
	gtttttccat	aggctccgcc	cccctgacga	gcatcacaaa	aatcgacgct	caagtccagag	2220
	gtggcgaaac	ccgacaggac	tataaagata	ccaggcgttt	ccccctggaa	gctccctcgt	2280
	gcgctctcct	gttcggaccc	tgccgcttac	cggataacctg	tcgcctttc	tccttcgggg	2340
	aagcgtggcg	ctttctcata	gctcacgctg	taggtatctc	agttcggtgt	aggtcggtcg	2400
30	ctccaagctg	ggctgtgtgc	acgaaccccc	cgttcagccc	gacgcgtcg	ccttatccgg	2460
	taactatcgt	cttgagttca	acccggtaag	acacgaacta	tcgccactgg	cagcagccac	2520
	tggtaacagg	attagcagag	cgaggatagt	aggcgtgtgt	acagagttct	tgaagtgggtg	2580
	gcctaactac	ggctacacta	gaaggacagt	atttggtatc	tgcgctctgc	tgaagccagt	2640
	taccttcgga	aaaagagttg	gtagctcttg	atccggcaaa	caaaccacog	ctggtagcgg	2700
35	tggttttttt	gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	2760
	tttgatcttt	tctacggggt	ctgacgctca	gtggaacgaa	aactcacggt	aagggatttt	2820
	ggtcagtga	ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagtgtt	2880
	taaataaatc	taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	2940
	tgaggcacct	atctcagoga	tctgtctatt	tcgttcaccc	atagttgctt	gactccccgt	3000
40	cgtgtagata	actacgatac	gggagggctt	accatctggc	cccagtgctg	caatgatacc	3060
	gcgagaccca	cgctcacccg	ctccagattt	atcagcaata	aaccagccag	ccggaaggggc	3120
	cgagcgcaga	agtggtoctg	caactttatc	cgcctccatc	cagtctatta	attgttgccg	3180
	ggaagctaga	gtaagttagt	cgccagttaa	tagtttgccg	aacgttggtg	ccattgctac	3240
	aggcatcgtg	gtgtcacgct	cgtcgtttgg	tatggcttca	ttcagctccg	gttcccaacg	3300
45	atcaaggcga	gttacatgat	cccccatggt	gtgcaaaaaa	gcggttagct	ccttcgggtc	3360
	tccgatcggt	gtcagaagta	agttggccgc	agtgttatca	ctcatgggtta	tggcagcaact	3420
	gcataattct	cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	3480
	aaccaagtca	ttctgagaat	agtgtatgag	gcgaccgagt	tgctcttgcc	cggcgtcaat	3540
	acggggata	accgcgccac	atagcagaac	tttaaaagtg	ctcatcattg	gaaaacggtc	3600
50	ttcggggcga	aaactctcaa	ggatcttacc	gttgttgaga	tccagttcga	tgtaaccacc	3660
	tcgtgcaccc	aactgatott	cagcatcttt	tactttcacc	agcgtttctg	ggtagagcaa	3720
	aacaggaagg	caaaatgccg	caaaaaaggg	aataagggcg	acaaggaaat	gttgaatact	3780
	catactcttc	ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	3840
	atacatattt	gaatgtattt	agaaaaataa	acaaataggg	gttccgcgca	catttccccg	3900
55	aaaagtgc						3908

<210> 71

<211> 3927

<212> DNA

60 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pBS-SSV4

65

<400> 71

cacctaaatt gtaagcggtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag 60

	ctcattttttt	aaccaatagg	ccgaaatcgg	caaaatccct	tataaatcaa	aagaatagac	120
	cgagatagg	ttgagtggtg	ttccagtttg	gaacaagagt	ccactattaa	agaacgtgga	180
	ctccaacgtc	aaagggcgaa	aaaccgtcta	tcagggcgat	ggccactac	gtgaaccatc	240
5	accctaataca	agtttttttg	ggtcgaggtg	ccgtaaagca	ctaaatcgga	accctaagg	300
	gagccccga	tttagagctt	gacggggaaa	gccggcgaa	gtggcgagaa	aggaaggga	360
	gaaagcgaaa	ggagcgggcg	ctagggcgct	ggcaagtgt	gcggtcacgc	tgcgcgtaac	420
	caccacaccc	gcccgcgtta	atgcccgcgt	acagggcgcg	tccatttcgc	cattcaggct	480
	gcgcaactgt	tgggaaggcg	gatcggtgcg	ggcctcttcg	ctattacgcc	agctggcgaa	540
	agggggatgt	gctgcaaggc	gattaagttg	ggtaacgcca	gggttttccc	agtcacgacg	600
10	ttgtaaaacg	acggccagtg	aattgtaata	cgactcacta	tagggcgaa	tggagctcca	660
	ccgcggtggc	ggcgcgacca	tgcccaagaa	gaagaggaag	gtgacgaaag	ataagacgcg	720
	ttataaatac	ggggattata	ttttacgcga	gaggaaaggg	cggattattg	tttacaagct	780
	agagtatgaa	aacgggtgag	taaaagagcg	ttacgtgggt	cctttagctg	acgtogttga	840
	atcatatcta	aaaatgaaat	taggggtcgt	aggggatact	cccctacaag	cggatcccc	900
15	cgggtttcgag	cccgggacaa	gcggaagcgg	tgggtgaaaa	gagggaaact	aacgacgtaa	960
	aatagcgttg	gttgccaatt	tgcgccaaata	cgcgacggac	ggcaacataa	aggcgttcta	1020
	caactatctc	atgaacgaaa	gggggataag	cgaaaaaact	gcaaaggact	acatcaatgc	1080
	tatatcaaa	ccgtataaag	agacgagaga	cgcacagaag	gcttaccgac	tctttgcacg	1140
	tttcttagcg	tcacgcaata	tcatacatga	tgaatttgcg	gataaaatat	tgaagcggt	1200
20	aaaggtgaag	aaggcgaaag	ctgatatcta	cattccaacg	ttggaagaga	taaaaaggac	1260
	gttacaatta	gcaaaagact	atagcgaaaa	cgtctacttc	atctaccgta	tgcgtctcga	1320
	gtcgggctgt	aggctgagcg	aaatactgaa	agtgcgtgaag	gaaccggaaa	gggacatttg	1380
	cggtaacgac	gtctgttatt	atccgcttag	ttggactagg	ggatataagg	gcgtcttcta	1440
	tgtattccac	ataacgcctc	tgaagagagt	agaggtgacg	aagtgggcaa	tagcggactt	1500
25	tgaacgacgt	cataaggacg	ctatagcgat	aaagtaactc	cgcaaatctc	tagcgtctaa	1560
	gatggctgag	ctaagcgtag	cgtagatat	tatcgctttt	attcaagggc	gtaaacggac	1620
	acgcgtttta	acgcaacatt	acgtatcgct	cttcggcata	gcgaaagagc	aataataaaaa	1680
	gtatgcggaa	tggctaaaag	gggtctgact	cgaggggggg	cccggtaacc	agcttttgtt	1740
	cccttttagtg	agggttaatt	tcgagcttgg	cgtaatcatg	gtcatagctg	tttctgtgt	1800
30	gaaattgtta	tccgctcaca	attccacaca	acatacgagc	cggaagcata	aagtgtaaa	1860
	cctggggtgc	ctaagtgtg	agctaactca	catttaattgc	gttgcgctca	ctgcccgtt	1920
	tccagtcggg	aaacctgtcg	tgccagctgc	attaatgaat	cggccaacgc	gcccgggagag	1980
	gcggttttgc	tattggggcg	tottccgctt	cctcgctcac	tgactcgctg	cgctcggtcg	2040
	ttcggctgcg	gcgagcggta	tcagctcact	caaaggcggt	aatacggtta	tccacagaat	2100
35	caggggataa	cgcaggaaa	aacatgtgag	caaaaggcca	gcaaaaggcc	aggaaccgta	2160
	aaaaggccgc	gttgcgtggc	ttttccata	ggctccgccc	ccctgacgag	catcacaaaa	2220
	atcgacgctc	aagtcagagg	tggcgaaacc	cgacaggact	ataaagatac	caggcgtttc	2280
	cccctggaag	ctccctcgta	cgctctcctg	ttccgacct	gccgcttacc	ggatacctgt	2340
	ccgcctttct	cccttcggga	agcgtggcgc	tttctcatag	ctcacgctgt	aggtatctca	2400
40	gttcggtgta	ggtcgttgc	tccaagctgg	gctgtgtgca	cgaaccccc	gttcagcccg	2460
	accgctgcgc	cttatccggt	aactatcgct	ttgagtccaa	cccggtaaga	cacgacttat	2520
	cgccactggc	agcacgact	ggtaacagga	ttagcagagc	gaggtatgta	ggcgggtgta	2580
	cagagttcct	gaagtgggtg	cctaactacg	gctacactag	aaggacagta	tttggttatc	2640
	gcgctctgct	gaagccagtt	accttcggaa	aaagagttgg	tagctcttga	tccggcaaac	2700
45	aaaccaccgc	tggtagcggt	ggtttttttg	tttgcaagca	gcagattacg	cgcagaaaaa	2760
	aaggatctca	agaagatcct	ttgatctttt	ctacggggtc	tgacgctcag	tggaaacgaaa	2820
	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	tagatccttt	2880
	taaattaaaa	atgaagtttt	aaatcaatct	aaagtatata	tgagttaaact	tggctgtaca	2940
	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	cggttcaccca	3000
50	tagttgcctg	actcccgcgc	gtgtagataa	ctacgatacg	ggagggttta	ccatctggcc	3060
	ccagtgcctg	aatgataccg	cgagacccac	gctcaccggc	tccagattta	tcagcaataa	3120
	accagccagc	cgggaaggcc	gagcgcagaa	gtggtcctgc	aactttatcc	gcctccatcc	3180
	agtctattaa	ttgttgccgg	gaagctagag	taagtagtcc	gccagttaat	agttttgcga	3240
	acgtttgttg	cattgctaca	ggcatcggtg	tgtaacgctc	gtcgttttgt	atggcttcat	3300
55	tcagctccgg	ttcccaacga	tcaaggcgag	ttacatgatc	ccccatgttg	tgcataaaag	3360
	cggttagctc	cttcggctct	cogacgttgg	tcagaagtaa	gttggccgca	gtgttatcac	3420
	tcattggttat	ggcagcactg	cataattctc	ttactgtcat	gccatccgta	agatgctttt	3480
	ctgtgactgg	tgagtactca	accaagtcat	tctgagaata	gtgtatgctg	cgaaccgagt	3540
	gctcttgccc	ggcgtcaata	cgggataata	ccgcgccaca	tagcagaact	ttaaaagtgc	3600
60	tcattcattgg	aaaacgttct	tcggggcgaa	aaactctcaag	gatcttaacc	ctgttgagat	3660
	ccagttcgat	gtaacccact	cgtgcaccca	actgatcttc	agcatctttt	actttcacca	3720
	gcgtttctgg	gtgagcaaaa	acaggaaggc	aaaatgccgc	aaaaaaggga	ataaggcgga	3780
	cacggaaatg	ttgaatactc	atactcttcc	tttttcaata	ttattgaagc	atattatcagg	3840
	gttattgtct	catgagcgga	tacatatttg	aatgtattta	gaaaaataaa	caaatagggg	3900
65	ttccgcgcac	atttcccgca	aaagtgc				3927

<210> 72
 <211> 3351
 <212> DNA
 <213> Artificial Sequence

5

<220>
 <223> Description of Artificial Sequence: vector
 pBS-SSVs

10

<400> 72
 cgacctcgag tcagaccctt tttagccatt ccgcatactt tttatattgc tctttcgota 60
 tgccgaagag cgatacgtaa tgttgcgtta aaacgcgtgt cggtttacgc ccttgaataa 120
 aatcgataat atctaaccgt acgottagct cagccatctt agacgctacg aatttgcgga 180
 agtactttat cgctatagcg tcttatgac gtcgttcaaa gtcogctatt gccacttcg 240
 15 tcacctctac tctcttcaga ggogttatgt ggaatacata gaagacgccc ttatatcccc 300
 tagtccaact aagcggataa taacagacgt cgttacgcga aatgtccctt tcgggttcct 360
 tcagcacttt cagtatttcg ctcagccata cgcgccgactc gagggggggc ccggtaccca 420
 gcttttgctt ccttttagtga gggtttaatt cgagcttgcc gtaatcatgg tcatagctgt 480
 ttctgtgtg aaattgttat ccgctcacaa cctacacaca catacgagcc ggaagcataa 540
 20 agtgtaaagc ctgggggtgccc taatgagtga gctaaactac attaatgctg ttgcgctcac 600
 tgcccgtttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgag 660
 cggggagagg cggtttgctg attggggcgt cttccgcttc ctcgctcact gactcgctgc 720
 gctcggctgt tcggctgcgg cgagcgggat cagctcactc aaaggcggta atacggttat 780
 ccacagaatc aggggataac gcaggaaga acatgtgagc aaaaggccag caaaaggcca 840
 25 ggaaaccgtaa aaaggccgct ttgctggcgt ttttccatag gctccgcccc cctgacgagc 900
 atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc 960
 aggcgttttc ccttgaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg 1020
 gatacctgtc cgcctttctc ccttggggaa gctgtggcgt ttctcatagc tcacgctgta 1080
 ggtatctcag ttccgtgtag gtctgtcgt ccaagctggg ctgtgtgcac gaaccccccg 1140
 30 ttacgcccga ccgctgcgcc ttatccgcta actatcgtct tgagtccaac ccggttaagc 1200
 acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag 1260
 gcggtgctac agagtctctg aagtgttgcc ctaactacgg ctacactaga aggacagtat 1320
 ttggtatctg cgtctcgtcg aagccagtta ccttcggaaa aagagttggt agctcttgat 1380
 ccggcaaaaa aaccaccgct ggtagcgggt gtttttttgt ttgcaagcag cagattacgc 1440
 35 gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt 1500
 ggaaacgaaaa ctcacgttaa gggatttttg tcatgagatt atcaaaaagg atcttcaact 1560
 agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaaact 1620
 ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc 1680
 gtccatccat agttgcctga ctcccgtcg tgtagataac tacgatacgg gagggcttac 1740
 40 catctggccc cagtgcctga atgataccgc gagaccacg ctcaccggct ccagatttat 1800
 cagcaataaa ccagccagcc ggaaggcccg agcgcagaag tggtcctgca actttatccg 1860
 cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttataa 1920
 gtttgccgca cgttggttgc attgtacag gcacgtggt gtcacgctcg tcgtttggt 1980
 45 tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt 2040
 gcaaaaaagc ggttagctcc ttoggtcctc cgatcggtgt cagaagttag ttggccgag 2100
 tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccacccgtaa 2160
 gatgcttttc tgttagtgt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc 2220
 gacgcagttg ctcttgccc gcgtcaatac gggataatac cgcgccacat agcagaactt 2280
 50 taaaagtgt catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc 2340
 tgttgagatc cagttcgatg taaccactc gtgcacccaa ctgatcttca gcatctttta 2400
 ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa 2460
 taaggcgac acggaaatgt tgaatactca tactcttctt ttttcaatat tattgaagca 2520
 55 tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac 2580
 aaataggggt tccgcgcaca tttcccggaa aagtgccacc taaattgtaa gcgttaatat 2640
 tttgttaaaa ttccggttaa atttttgtta aatcagctca ttttttaacc aataggccga 2700
 aatcggcaaa atcccttata aatcaaaaaga atagaccgag atagggttga gtgttgttcc 2760
 agtttggaac aagagtcacac tattaagaaa cgtggactcc aacgtcaaag ggcgaaaaac 2820
 cgtctatcac ggcgatggcc cactacgtga accatcccc taatcaagtt ttttggggtc 2880
 60 gaggtgccgt aaagcactaa atcggaaacc taataggagc ccccgattta gagcttgacg 2940
 gggaaaagcc gcgaacgtgg ccgaaaagga agggaaagaa gcgaaaaggag cggcgctag 3000
 ggcgctggca agtgtagcgg tcaogctgog cgtaaccacc acaccgcgag cgcttaatgc 3060
 gccgctacag ggcgcgtccc attcgcatt caggctgcgc aactgttggg aagggcgatc 3120
 ggtgcgggcc tcttcgctat taogccagct ggcgaaaagg ggatgtgctg caaggcgatt 3180
 65 aagttgggta acgcaggggt tttccagtc acgacgttgt aaaacgacgg ccagtgaatt 3240
 gtaatacgac tcactatagg gcgaattgga gctccaccgc ggtggcgcc gctctagaac 3300
 tagtggatcc cccgggctgc aggaattcga tatcaagctt atcgataaccg t 3351

<210> 73

<211> 5730

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pCMVC31 (NNLS)

<400> 73

aaacagtcog	atgtacgggc	cagatatagc	cgttgacatt	gattattgac	tagttattaa	60
tagtaatacaa	ttacgggggc	attagttcat	agcccatata	tggagttccg	cgttacataa	120
cttacgggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccgccatt	gacgtcaata	180
atgacgtatg	ttcccatagt	aacgccataa	gggactttcc	attgacgtca	atgggtggac	240
tattttacggt	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	aagtacgcc	300
cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	catgacctta	360
tgggactttc	ctaactggca	gtacatctac	gtattagtoa	tgcctattac	catggtgatg	420
cgggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	atttccaagt	480
ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	540
aaatgtcgta	acaactccgc	cccatggacg	caaatgggog	gtaggcgtgt	acgggtgggag	600
gtctatataa	gcagagctct	ctggctaact	agagaaccca	ctgcttactg	gcttatcgaa	660
attaatacga	ctcaactatag	ggagacccaa	gctgactcta	gacttaatta	agcgttgggg	720
tgagtaactcc	ctctcaaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	780
cgaggaggat	ttgatattca	cctggccgcg	ggtgatgcct	ttgagggtgg	ccgcgtccat	840
ctggtcagaa	aagacaatct	ttttgttgtc	aagcttgagg	tgtggcaggc	ttgagatctg	900
gccatacaact	tgagtgcacat	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	960
cagggcgggcc	gcaccatgcc	caagaagaag	aggaaggtga	cacaaggggt	tgtgaccggg	1020
gtggacacgt	acgcgggtgc	ttacgaccgt	cagtcgcgcg	agcgcgagaa	ttcagacgca	1080
gcaagcccag	cgacacagcg	tagcgccaac	gaagacaagg	cggccgacct	tcagcgcgaa	1140
gtcagacggg	acgggggccc	gttcaggttc	gtcgggcatt	tcagcgaagc	gccgggcacg	1200
tccggcttcg	ggacggcgga	gcgcccggag	ttcgaaacga	tcctgaacga	atgcccgccg	1260
gggcggtcca	acatgatcat	tgtctatgac	gtgtgcgcgt	tctcgcgcct	gaaggtcatg	1320
gacgcgattc	cgattgtctc	ggaattgtct	gccctggggc	tgacgattgt	ttccactcag	1380
gaaggcgtct	tcgggcaggg	aaacgtcatg	gacctgattc	acctgattat	gcggctcgac	1440
gcgtcgacac	aagaattctt	gctgaagtcg	gcgaagattc	tcgacacgaa	gaaccttcag	1500
cgcgcaattgg	cgcggtacgt	cggcggaag	gccttcacg	gcttcgagct	tgtttcgag	1560
acgaaggaga	tcacgcgcaa	cggccgaatg	gtcaatgtcg	tcatacaaaa	gcttgcgcac	1620
tcgaccactc	cccttaacgg	acccttcgag	ttcgagccc	acgtaatccg	gtgggtgggtg	1680
cgtgagatca	agacgcacaa	acaccttccc	ttcaagccgg	gcagtcagc	cgccattcac	1740
ccgggcagca	tcacggggct	ttgtaagcgc	atggacgctg	acgcccgtgc	gacccggggc	1800
gagacgattg	ggaagaagac	cgcttcaagc	gcttgggacc	cggcaaccgt	tatgcgaatc	1860
cttcgggacc	cggcttttgc	ggccttcgcc	gctgaggtga	tctacaagaa	gaagccggac	1920
ggcacgcgga	ccacgaagat	tgagggttac	cgcattcagc	gcgaccgat	cacgctccgg	1980
ccggtcgagc	ttgattgcgg	accgatcatc	gagcccgtcg	agtggatatga	gcttcaggcg	2040
tggttggacg	gcagggggcg	cggcaagggg	ctttcccggg	ggcaagccat	tctgtccgcc	2100
atggacaagc	tgtactgcga	gtgtggcgcc	gtcatgactt	cgaagcgcgg	ggaagaatcg	2160
atcaaggact	cttaccgctg	ccgtcgcgg	aagcttgctg	acccgtccgc	acctggcgag	2220
cacgaaggca	cgtgcaacgt	cagcatggcg	gcactgcaca	agttcgttgc	ggaacgcac	2280
ttcaacaaga	tcaggcacgc	cgaaggcgac	gaagagacgt	tggcgcttct	gtgggaagcc	2340
gcccagacgt	tcggcaagct	cactgaggcg	cctgagaaga	gcggcgaaac	ggcgaaacct	2400
gttgccggagc	gcgcgcagcg	cctgaacgcc	cttgaagagc	tgtacgaaga	ccgcgcggca	2460
ggcgcgtagc	acggacccgt	tggcaggaag	cacttccgga	agcaacaggc	agcgtcgacg	2520
ctccggcagc	aaggggcgga	agagcggctt	gccgaacttg	aagccgcgga	agcccgaag	2580
cttccccttg	accaatgggt	ccccgaagac	ccgcagcgtg	acccgaccgg	ccctaagtcg	2640
tgggtgggggc	gcgcgtcagt	agacgacaag	cgcggtgttc	tcgggctctt	cgtagacaag	2700
atcgttgtca	cgaagtcgac	tacgggcagg	gggcagggaa	cgcccatcga	gaagcgcgct	2760
tcgatcacgt	ggcggaagcc	gccgaccgac	gacgacgaag	acgacgcca	ggacggcacg	2820
gaagacgtag	cggcgtaggc	ggcgcccggg	ctcgagatcc	aggcgcggat	caataaaaaga	2880
tcattattttt	caatagatct	gtgtgttgg	ttttgtgtg	ccttggggga	gggggagggc	2940
agaatgaggc	gcggccaagg	gggaggggga	ggccagaatg	accttggggg	agggggaggc	3000
cagaatgacc	ttgggggagg	gggagggcag	aatgaggcgc	gccccggggt	accgagctcg	3060
aattcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	tacccaactt	3120
aatcgccctt	cagcacatcc	ccctttccgc	agctggcgta	atagcgaaga	ggccgcgacc	3180
gatccgccct	cccaacagtt	gcgcagcctt	aatggcgaat	ggcgccctgat	gcggtatttt	3240
ctccttacgc	tattgtgcgg	tatttcacac	cgcattatgt	gcactctcag	tacaatctgc	3300
tctgatgcgc	catagtttaag	ccagccccga	cacccgcca	caccgctga	cgcgccctga	3360

	cgggcttgtc	tgctcccggc	atccgcttac	agacaagctg	tgacogtctc	cgggagctgc	3420
	atgtgtcaga	ggttttcacc	gtcatcaccg	aaacgcgcga	gacgaaaggg	cctcgtgata	3480
	cgccatattt	tataggttaa	tgcatgata	ataatggttt	cttagacgtc	aggtggcact	3540
5	tttoggggaa	atgtgcgcgg	aacccctatt	tgtttatttt	tctaaataca	ttcaaataatg	3600
	tatccgctca	tgagacaata	accctgataa	atgcttcaat	aatattgaaa	aaggaagagt	3660
	atgagtattc	aacattttccg	tgctgcacct	attccctttt	ttgcggcatt	ttgccttccct	3720
	gtttttgtc	acccagaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	gttgggtgca	3780
	cgagtgggtt	acatcgaaat	ggatctcaac	agcggtaaga	tccttgagag	ttttcgcccc	3840
10	gaagaacgtt	ttccaatgat	gagcactttt	aaagtctctg	tatgtggcgc	ggtatttatcc	3900
	cgtattgacg	ccgggcaaga	gcaactcggg	cgccgcatac	actattctca	gaatgacttg	3960
	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcattgacagt	aagagaatta	4020
	tgcatgtctg	ccataaccat	gagtataaac	actgcggcca	acttaactct	gacaacgatc	4080
	ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	aaactcgccct	4140
15	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	caccacgatg	4200
	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	tactctagct	4260
	tcccggaac	aattaataga	ctggatggag	gcgataaaag	ttgcaggacc	acttctgcgc	4320
	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccgggtg	gcgtgggtct	4380
	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	ccogtatcgt	agttatctac	4440
20	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	gatagggtgc	4500
	tcactgatta	agcattggta	actgtcagac	caagtttact	catatatact	ttagattgat	4560
	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	tcctttttga	taactctcatg	4620
	acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	cagaccccg	agaaaagatc	4680
	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaactc	gctgcttgca	aacaaaaaaa	4740
25	ccaccgctac	cagcgggtgg	ttgtttgcgg	gatcaagagc	taccaactct	ttttccgaag	4800
	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgtg	gccgtagtta	4860
	ggccaccact	tcaagaacte	tgtagcaccg	cctacatacc	tcgctctgct	aatcctgtta	4920
	ccagtggctg	ctgccagtgg	cgataagtcg	tgtcttaccg	ggttggactc	aagacgatag	4980
	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggt	cgtgcacaca	gcccagcttg	5040
30	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	agctatgaga	aagcgccacg	5100
	cttcccggaag	ggagaaaggc	ggacaggat	ccggtaagcg	gcagggtcgg	aacaggagag	5160
	cgcacgaggg	agcttccagg	gggaaacgcc	tggtatcttt	atagtcctgt	cgggtttcgc	5220
	cacctctgac	ttgagcgtcg	attttttgtg	tgctcgtcag	gggggcggag	cctatggaaa	5280
	aacgccagca	acgcgcctt	tttacggctt	ctggcctttt	gctggccttt	tgctcacatg	5340
35	ttctttcctg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	tgagttagct	5400
	gataccgctc	gcgcgagccg	aacgacccag	cgcagcaggt	cagttagcga	ggaagcggaa	5460
	gagcgcccaa	taogcaaac	gcctctcccc	gcgcgttggc	cgattcatta	atgcagctgg	5520
	cacgacaggt	ttcccgactg	gaaagcgggc	agttagcgcga	acgcaattaa	tgtgagttag	5580
	ctcactcatt	aggcacccca	ggctttacac	tttatgcttc	cggctcgtat	gttggttggg	5640
40	attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	cgccaagcta	5700
	gcccgggcta	gcttgcatgc	ctgcagggtt				5730

<210> 74

<211> 4886

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector pCMV-SSV

<400> 74

	aaacagtccg	atgtacgggc	cagatatacg	cgttgacatt	gattattgac	tagttattaa	60
	tagtaataca	ttacgggggc	attagttcat	agcccatata	tggagttccg	cgttacataa	120
55	cttacggtaa	atggcccgcg	tggtgacccg	cccaacgacc	cccgcctatt	gacgtcaata	180
	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	atgggtggac	240
	tattttacgg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	aagtagcccc	300
	cctatttgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	catgacctta	360
	tgggaactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	catggtgatg	420
60	cgggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	atttccaagt	480
	ctccaccoca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	540
	aaatgtcgtg	acaactccgc	cccattgacg	caaattggcg	gtaggcgtgt	acgggtgggag	600
	gtctatataa	gcagagctct	ctggctaact	agagaaccca	ctgcttactg	gcttatcgaa	660
	attaatacga	ctcactatag	ggagacccaa	gctgactcta	gacttaatta	agcgttgggg	720
65	tgagtactcc	ctctcaaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	780
	cgaggaggat	ttgatattca	cctggcccgc	ggtgatgcct	ttgagggtgg	ccgcgtccat	840
	ctggtcagaa	aagacaatct	ttttgttctc	aagcttgagg	tgtggcaggc	ttgagatctg	900

	gccatacact	tgagtgcacat	tgacatccac	tttgcctttt	tctccacagc	tgtccactcc	960
	cagggcgccc	gcccgatatg	acgaaagata	agacgogtta	taaatacggg	gattatatit	1020
	tacgcgagag	gaaagggcgg	tattatgttt	acaagctaga	gtatgaaaac	ggtgaggtaa	1080
5	aagagcggtta	cgtgggtcct	ttagctgacg	togttgaatc	atatctaaaa	atgaaattag	1140
	gggtcgtagg	ggatactccc	ctacaagcgg	atccccccgg	tttcgagccc	gggacaagcg	1200
	gaagcgggtg	tggaaaagag	ggaactgaac	gacgtaaaaat	agcgttggtt	gccaatttgc	1260
	gccaatacgc	gacggacggc	aacataaagg	cgttctacaa	ctatctcatg	aacgaaaggg	1320
	ggataagcga	aaaaactgca	aaggactaca	tcaatgctat	atcaaagccg	tataaagaga	1380
10	cgagagacgc	acagaaggct	taccgactct	ttgcaogttt	cttagcgtca	cgcaatatca	1440
	tacatgatga	atttgcggat	aaaatattga	aagcggtaaa	ggtgaagaag	gcgaacgctg	1500
	atatctacat	tccaacggtt	gaagagataa	aaaggacggt	acaattagca	aaagactata	1560
	gcgaaaaagt	ctacttcctc	taccgtatcg	ctctcgagtc	gggcggttag	ctgagcgaaa	1620
	tactgaaagt	gctgaaggaa	cccgaaaggg	acattttgogg	taacgacgct	tgttattatc	1680
15	cgcttagttg	gactagggga	tataaggcgg	tcttctatgt	attccacata	acgcctctga	1740
	agagagtaga	ggtgacgaag	tgggcaatag	cggactttga	acgacgctcat	aaggacgcta	1800
	tagcgataaa	gtacttcggc	aaattcgtag	cgtctaagat	ggctgagcta	agcgtaccgt	1860
	tagatattat	cgatttttatt	caagggcgta	aaccgacacg	cgtttttaacg	caacattacg	1920
	tatcgctctt	cggcatagcg	aaagagcaat	ataaaaaagta	tgcggaatgg	ctaaaagggg	1980
20	tctgactcga	ggggggggccc	gtcgacctcg	agatccaggc	gcggatcaat	aaaagatcat	2040
	tattttcaat	agatctgtgt	gttgggtttt	tgtgtgcctt	gggggagggg	gaggccagaa	2100
	tgaggcgccg	ccaaggggga	gggggaggcc	agattgacct	tgggggaggg	ggagccagaa	2160
	atgaccttgg	gggaggggga	ggccagaatg	aggcgcgccc	ccgggtaccg	agctcgaatt	2220
	cactggccgt	cgttttacaa	cgctcgtgact	gggaaaaacc	tggcggtacc	caacttaatc	2280
25	gccttgcagc	acatccccct	ttcgccagct	ggcgtaatat	cgaagaggcc	cgcaccgatc	2340
	gcccctccca	acagttgcgc	agcctgaatg	gcgaattggcg	cctgatgcgg	tattttctcc	2400
	ttacgcctct	gtgcggtatt	tcacaccgca	tatggtgcac	tctcagtaaa	atctgctctg	2460
	atgcgcata	gttaagccag	ccccgacacc	cgcccaacacc	cgctgacgcg	ccctgacggg	2520
	cttgtctgct	cccggcatcc	gcttacagac	aagctgtgac	cgtctccggg	agctgcatgt	2580
30	gtcagaggtt	ttcacccgtca	tcaccgaaac	gcgcgagacg	aaagggcctc	gtgatacgcc	2640
	tatttttata	ggttaatgtc	atgataataa	tggtttctta	gacgtcaggt	ggcacttttc	2700
	ggggaaatgt	gcgcggaacc	cctattttgtt	tatttttcta	aatacattca	aatatgtatc	2760
	cgctcatgag	acaataaccc	tgataaatgc	ttcaataata	ttgaaaaagg	aagagtatga	2820
	gtattcaaca	tttccgtgtc	gccccttatcc	ccttttttgc	ggcattttgc	cttccgtgtt	2880
35	ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	ggtgcacgag	2940
	tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	3000
	aacgttttcc	aatgatgagc	acttttaaa	ttctgctatg	tggcgcggta	ttatccccta	3060
	ttgaocgcgg	gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	gacttggttg	3120
	agtaactcac	agtcacagaa	aagcatctta	cgtatggcat	gacagtaaga	gaattatgca	3180
40	gtgctgccat	aacctatgag	gataacactg	cggccaactt	acttctgaca	acgatcggag	3240
	gaccgaagga	gctaaccgct	tttttgcaaa	acatggggga	tcatgttaact	cgccttgatc	3300
	ggttggaacc	ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	acgatgcctg	3360
	tagcaaatgg	aacaacgttg	cgcaaaacta	taactggcga	actacttact	ctagcttccc	3420
45	ggcaacaatt	aatagactgg	atggaggcgg	ataaagttgc	aggaccactt	ctgcgctcgg	3480
	cccttccggc	tggctggttt	attgctgata	aatctggagc	cgggtgagcgt	gggtctcggc	3540
	gtatcattgc	agcactgggg	ccagatggta	agccctcccg	tatcgtagtt	atctacacga	3600
	cggggagtcg	ggcaactatg	gatgaacgaa	atagacagat	cgtcgagata	ggtgcctcac	3660
	tgattaagca	ttggttaactg	tcagaccaag	tttactcata	tatacttttag	attgatttaa	3720
	aacttcattt	ttaatttaaa	aggatctagg	tgaagatcct	tttttgataat	ctcatgacca	3780
50	aaatccctta	acgtgagttt	tgcgttccact	gagcgtcaga	cccogtagaa	aagatcaaa	3840
	gatcttcttg	agatcccttt	tttctgcgcg	taatctgctg	ottgcaaaca	aaaaaaccac	3900
	cgctaccagc	ggtggtttgt	ttgcgggac	aagagctacc	aaotcttttt	ccgaaggtaa	3960
	ctggcttcag	cagagcgag	ataccaaata	ctgtccttct	agtgtagccg	tagttaggcc	4020
	accacttcaa	gaactctgta	gcaccgccta	catacctcgc	tctgctaato	ctgttaccag	4080
55	tggctgctgc	cagtgctgat	aagtcgctgc	ttaccgggtt	ggactcaaga	cgatagttac	4140
	cggataaagg	gcagcggtcg	ggctgaacgc	ggggttcgtg	cacacagccc	agcttgagc	4200
	gaacgacctg	caccgaactg	agatacctac	agcgtgagct	atgagaaagc	gccacgcttc	4260
	ccgaaggag	aaaggcggac	aggtatccgg	taagcggcag	ggtcgggaaca	ggagagcgca	4320
	cgagggagct	tccaggggga	aacgcctggt	atctttatag	tctgtcggg	tttcgccacc	4380
60	tctgacttga	gcgtcgattt	ttgtgatgct	cgtcaggggg	gcggagccta	tggaaaaacg	4440
	ccagcaacgc	ggcctttttt	cggttcctgg	ccttttctg	gccttttgc	cacatgttct	4500
	ttcctcgctt	atccccgat	tctgtggata	accgtattac	cgcctttgag	tgagctgata	4560
	ccgctcgccg	cagccgaacg	accgagcgca	gcgagtcagt	gagcgaggaa	gcggaagagc	4620
	gccccatacg	caaaccgcct	ctccccgcgc	gttgccgat	tcattaatgc	agctggcacg	4680
65	acaggtttcc	cgactggaaa	gcgggcagtg	agcgcaacgc	aattaatgtg	agttagctca	4740
	ctcattagga	accccagget	ttacacttta	tgcttccggc	togtatgttg	tgtggaattg	4800
	tgagcggata	acaatttcac	acaggaacaa	gctatgacca	tgattacgcc	aagctagccc	4860
	gggctagctt	gcattgcctgc	aggttt				4886

<210> 75
 <211> 4905
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: vector
 pCMV-SSV(NNLS)

<400> 75
 aaacagtcgg atgtacgggc cagatatagc cgttgacatt gattattgac tagttattaa 60
 tagtaatacaa ttacgggggc attagttcat agcccatata tggagttccg cgttacataa 120
 cttacgggtaa atggcccgcc tggctgaccg cccaacgacc cccgccatt gacgtcaata 180
 atgacgtatg ttcccatagt aacgccataa gggactttcc attgacgtca atgggtggac 240
 tattttacgg aaactgccc aattggcagta catcaagtgt atcatatgcc aagtacgccc 300
 cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta catgacctta 360
 tgggaacttt ctaacttgga gtacatctac gtattagtca tcgctattac catgggtgatg 420
 cggtttttgg agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
 ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggaactttcca 540
 aaatgtcgta acaactccgc cccattgacg caaatggcg gtaggcgtgt acgggtgggag 600
 gtctatataa gcagagctct ctggctaact agagaaccca ctgcttactg gcttatcgaa 660
 attaatacga ctcaactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
 tgagtactcc ctctcaaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa 780
 cgaggaggat ttgatattca cctggcccg cgtgatgcct ttgaggtgg cgcgctccat 840
 ctggtcagaa aagacaatct ttttgttgtc aagcgtgagg tgtggcaggc ttgagatctg 900
 gccatacaact tgagtacat tgacatccac tttgccttcc tctccacagg tgtccactcc 960
 cagggcgggc gcaccatgcc caagaagaag aggaaggtga cgaaagataa gacgcgttat 1020
 aaatacgggg attatatatt acgcgagagg aaagggcgg attatgttta caagctagag 1080
 tatgaaaaac gtgaggtaaa agagcgttac gtgggtcctt tagctgacgt cgttgaatca 1140
 tatctaaaaa tgaaattagg gtcgttaggg gatactcccc tacaagcggg tccccccgg 1200
 ttccagcccg ggacaagcgg aagcgggtgg ggaaaagagg gaactgaacg acgtaaaaa 1260
 gcgttgggtg ccaatttggc ccaatacgcg acggacggca acataaaggc gttctacaac 1320
 tatctcatga acgaaaggg gataagcgaa aaaactgcaa aggactacat caatgctata 1380
 tcaaagccgt ataaagagac gagagacgca cagaaggctt accgactctt tgcaactttc 1440
 ttacgctcac gcaattatcat acatgatgaa tttgcggata aaatattgaa agcggtaaa 1500
 gtgaagaagg cgaacgtgta tatctacatt ccaacgttgg aagagataaa aaggacgtta 1560
 caattagcaa aagactatag cgaaaacgtc tacttcatct accgtatcgc tctcgagtcg 1620
 ggcgttaggc tgagcgaaat actgaaagt ctgaaggaa cggaaaggga catttgcgg 1680
 aacgacgtct gttattatoc gcttagttgg actaggggat ataagggcgt cttctatgta 1740
 ttccacataa cgcctctgaa gagagtagag gtgacgaagt gggcaatagc ggactttgaa 1800
 cgacgtcata aggcgctat agcgataaag tacttccgca aattcgtagc gtctaagatg 1860
 gctgagctaa gcgtaccgtt agatattatc gattttatcc aagggcgtaa accgacacgc 1920
 gttttaacgc aacattacgt atcgtctctc ggcatagcga aagagcaata taaaaagtat 1980
 gcggaatggc taaaaggggt ctgactcgag ggggggcccg tcgacctcga gatccaggcg 2040
 cggatcaata aaagatcatt attttcaata gatctgtgtg ttggtttttt gtgtgccttg 2100
 ggggaggggg agggccgaat gaggcgcggc caagggggag ggggaggcca gaatgacctt 2160
 gggggagggg gagggcagaa tgaccttggg ggagggggag gccagaatga ggcgcgccc 2220
 cgggtaccga gctcgaattc actggccgtc tttttacaac gtcgtgactg ggaaaaacct 2280
 ggcgttaacc aacttaatcg ccttgacgca catccccctt tcgccagctg gcgtaatagc 2340
 gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatgggc 2400
 ctgatgcgtt attttctcct tacgcactcg tcgggtatit cacaccgat atgggtgact 2460
 ctcatgacaa tctgctctga tgccgcatg ttaagccagc cccgacaccc gccaacaccc 2520
 gctgacgcgc cctgacgggc ttgtctgctc ccggcatccg cttacagaca agctgtgacc 2580
 gtctccggga gctgcatgtg tcagagggtt tcaccgtcat caccgaaacg cgcgtagcga 2640
 aagggcctcg tgatacgctt atttttatag gttaatgtca tgataataat ggtttcttag 2700
 acgtcaggtg gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa 2760
 atacattcaa atatgtatcc gctcatgaga caataacct gataaatgct tcaataatat 2820
 tgaaaaagga agagtattga tattcaacat ttccgtgtcg ccttatttcc cttttttgag 2880
 gctatttggc ttctgttttt tgctcaccca gaaacgtctg tgaaagttaa agatgtgaa 2940
 gatcagttgg gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt 3000
 gagagttttc gccccgaaga acgtttttcca atgatgagca cttttaaagt tctgctatgt 3060
 ggcgcggtat tatcccgat tgacgcgggg caagagcaac tcggtcgccg catacactat 3120
 tctcagaatg acttggttga gtactcacca gtacagaaa agcatcttac ggatggcatg 3180
 acagtaagag aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta 3240
 cttctgacaa cgtatcgagg accgaaggag ctaaccgctt ttttgcacaa catgggggat 3300

65

	catgtaactc	gccttgatcg	ttgggaaccg	gagctgaatg	aagccatacc	aaacgacgag	3360
	cgtgacacca	cgatgcctgt	agcaatggca	acaacgttgc	gcaaaactatt	aactggcgaa	3420
	ctacttactc	tagcttcccg	gcaacaatta	atagactgga	tggaggcgga	taaagttgca	3480
5	ggaccacttc	tgcgctcggc	ccttcoggt	ggctgggtta	ttgctgataa	atctggagcc	3540
	ggtagcgtg	ggctcgcgg	tatcattgca	gcactggggc	cagatggtaa	gccctcccgt	3600
	atcgtagtta	tctacacgac	ggggagtcag	gcaactatgg	atgaacgaaa	tagacagatc	3660
	gctgagatag	gtgcctcaact	gattaagcat	tggtaactgt	cagaccaagt	ttactcatat	3720
	atactttaga	ttgattttaa	acttcatttt	taattttaaaa	ggatctaggt	gaagatcctt	3780
10	tttgataatc	tcatgaccaa	aatoccttaa	cgtgagtttt	cgttccactg	agcgtcagac	3840
	cccgtagaaa	agatcaaagg	atctttctga	gatccttttt	ttctgcgcgt	aatctgctgc	3900
	ttgcaaacia	aaaaaccacc	gotaccagcg	gtgggtttgt	tgccggatca	agagctacca	3960
	actctttttc	cgaaggtaac	tggcttcagc	agagcgcaga	taccaaatac	tgtccttcta	4020
	gtgtagccgt	agttaggcca	ccacttcaag	aactctgtag	caccgcctac	atacctcgct	4080
15	ctgctaattc	tgttaccagt	ggctgctgoc	agtggcgata	agtctgtct	taccgggttg	4140
	gactcaagac	gatagttacc	ggataaggcg	cagcggtcgg	gctgaacggg	gggttcgtgc	4200
	acacagccca	gcttggagcg	aacgacctac	accgaactga	gatacctaca	gcgtgagcta	4260
	tgagaaagcg	ccacgccttc	cgaagggaga	aaggcggaca	ggtatccggt	aagcggcagg	4320
	gtcggaaacg	gagagcgcac	gagggagcct	ccagggggaa	acgcctggta	tctttatagt	4380
20	cctgtcgggt	ttcgccacct	ctgacttgag	cgtcgatttt	tgtgatgctc	gtcagggggg	4440
	cgagccttat	ggaaaaacgc	cagcaacgcg	gcctttttac	ggttctctgc	cttttctgg	4500
	ccttttgctc	acatgttctt	tctcggtta	tcccttgatt	ctgtggataa	ccgtattacc	4560
	gcctttgagt	gagctgatac	cgctcgccgc	agccgaacga	ccgagcgcag	cgagtcagtg	4620
	agcgagggaag	cggaagagcg	cccaatacgc	aaacgcctc	tcccgcgcg	ttggccgatt	4680
25	cattaatgca	gctggcacga	caggtttccc	gactggaaag	cgggcagtga	gcgcaacgca	4740
	attaatgtga	gttagctcac	tcattaggga	ccccagcctt	tacactttat	gcttccggct	4800
	cgtatgttg	gtggaattgt	gagcggataa	caatttcaca	caggaaacag	ctatgaccat	4860
	gattacgcca	agctagcccg	ggctagcttg	catgcctgca	ggttt		4905
30	<210> 76						
	<211> 5290						
	<212> DNA						
	<213> Artificial Sequence						
35	<220>						
	<223> Description of Artificial Sequence: vector						
	pCMVXisA						
	<400> 76						
40	agtccgatgt	acggggccaga	tatacgcggt	gacattgatt	attgactagt	tattaatagt	60
	aatcaattac	ggggtcatta	gttcatagcc	catatatgga	gttcgcgcgt	acataactta	120
	cggtaaatgg	ccgccttggc	tgaccgcccc	acgacccccg	cccattgacg	tcaataatga	180
	cgtatgcttc	catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggactatt	240
45	tacggtaaac	tgcccacttg	gcagtagatc	aagtgtatca	tatgccaaat	acgcccccta	300
	ttgacgtcaa	tgacggtaaa	tggcccgctt	ggcattatgc	ccagtacatg	accttatggg	360
	actttcctac	ttggcagtac	atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggt	420
	tttggcagta	catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	480
	accccattga	cgtcaatggg	agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	540
	gtcgttaaca	ctccgcacca	ttgacgcaaa	tggccggtag	gcgtgtacgg	tgggaggtct	600
50	atataagcag	agctctctgg	ctaactagag	aacccactgc	ttactggcct	atcgaaatta	660
	atacgactca	ctatagggag	acccaagctg	actctagact	taattaagcg	ttgggggtgag	720
	tactccctct	caaaagcggg	catgacttct	gcgctaagat	tgtcagtttc	caaaaacgag	780
	gaggatttga	tattcacctg	gcccgcgggt	atgcctttga	gggtggccgc	gtccatctgg	840
	tcagaaaaga	caatcttttt	gttgtcaagc	ttgaggtgtg	gcaggcttga	gatctggcca	900
55	tacacttgag	tgacattgac	atccactttg	cctttctctc	cacaggtgtc	cactcccagg	960
	gcggccgccc	gatatgcaaa	atcagggtca	agacaaatat	caacaagcct	ttgcagactt	1020
	agagccactc	tcactctacc	acggcagttt	tctcggtcca	agtctgcaag	cacagcagca	1080
	aagagaacac	atgagaacaa	aagtactaca	agacctagac	aaggtaaatc	tgcgtttgaa	1140
60	gtctgcaaag	acgaaagtct	cagttcgaga	atctaaccgga	agtctgcaat	tacgagcaac	1200
	gttaccatatt	aaacctggag	ataaggacac	caacggtaca	ggcagaaagc	aatacaatct	1260
	cagcttgaaat	atccctgcaa	acttgatgg	actgaagacg	gctgaggaag	aagcttatga	1320
	attaggtaaa	ttaatcgctc	ggaaaacctt	tgaatggaat	gataaatatt	taggcaaaga	1380
	agccactaaa	aaagattcac	aaacaatagg	tgatttacta	gaaaaatttg	cagaagagta	1440
	ttttaaaacc	cataaacgca	ccactaaaag	cgaacatacc	tttttttact	atttttcccg	1500
65	cacccaacga	tataccaatt	ccaaagattt	agcaacggcg	gaaaatctca	tcaattcaat	1560
	tgagcaaatc	gataaagaat	gggcgagata	taatgccgc	agagccatat	cagctttttg	1620
	cataacattc	aatatagaaa	ttgatttgtc	ccagtattcc	aaaatgcctg	atcgcaattc	1680

	gcgcaacatc	cccacagatg	cagaaatact	atcaggaatt	accaaatttg	aagactatct	1740
	agttaccaga	ggaaatcaag	ttaatgaaga	tgtaaaagat	agctggcaac	tttggcgctg	1800
	gacatatgga	atgttagcag	tttttggttt	acgccccagg	gaaattttta	tttaacctaa	1860
5	tattgattgg	tggttaagca	aagagaatat	agacctcaca	tggaaagtag	acaaagaatg	1920
	taaaactggt	gaaagacaag	cattaccctt	acataaagaa	tggattgatg	agtttgattt	1980
	aagaaatccg	aaatattttag	aaatgctggc	aacagcaatt	agtaaaaaag	ataaaaacaaa	2040
	tcatgctgaa	ataacagcct	taactcagcg	tattagtttg	tggtttcgga	aagtcgaatt	2100
	agattttaaa	ccctatgatt	taoctcacgc	ctgggcaatt	agagcgcata	tttttaggcat	2160
10	accaatcaaa	gcggcggtcg	ataatttggg	gcatagtatg	caggttcata	cacaaaccta	2220
	tcagcgctgg	ttctcgctag	atatgoggaa	gttagcgatt	aatcaggctt	tgactaagag	2280
	gaatgaattt	gaggtgatta	gggaggagaa	tgctaaattg	cagatagaaa	atgaaagggt	2340
	gaggatggaa	attgagaagt	taaagatgga	aatagcttat	aagaatagtt	gagcggccgc	2400
	gtcgacctcg	agatccaggg	gcggatcaat	aaaagatcat	tattttcaat	agatctgtgt	2460
15	gthtggtttt	tgtgtgcctt	ggggggaggg	gaggccagaa	tgaggcgccg	ccaagggggg	2520
	gggggagggc	agaatgacct	tgggggaggg	gaggccagaa	atgaccttgg	gggaggggga	2580
	ggccagaatg	aggcgccgcc	ccgggtaccg	agctcgaatt	cactggccgt	cgttttacaa	2640
	cgctcgtgact	gggaaaaccc	tggcggtacc	caacttaatc	gccttgcagc	acatccccct	2700
	ttcgccagct	ggcgtaatag	cgaagaggcc	cgcaccgatc	gccttcccca	acagttgcgc	2760
20	agcctgaatg	gcgaattggc	cctgatgcgg	tattttctcc	ttacgcctct	gtgcgggtatt	2820
	tcacacgcga	ttcaatgcac	tctcagtcac	atctgctctg	atgcgcata	gttaagccag	2880
	ccccgacacc	cgccaacacc	cgctgacgcg	coctgacggg	cttgtctgct	ccgggcatcc	2940
	gcttacagac	aagctgtgac	cgtctccggg	agctgcattg	gtcagagggt	ttcaccgtca	3000
	tcaccgaaac	gcgcgagacg	aaagggcctc	gtgatacgcc	tatttttata	ggttaatgtc	3060
25	atgataataa	tggtttctta	gacgtcaggt	ggcacttttc	ggggaaatgt	gcgcgggaacc	3120
	cctattttgt	tattttttcta	aatacattca	aatatgtatc	cgctcatgag	acaataaccc	3180
	tgataaatgc	ttcaataata	ttgaaaaagg	aagagtatga	gtattcaaca	tttcogtgc	3240
	gcccattatc	ccttttttgc	ggcattttgc	cttcctgttt	ttgctcacc	agaaacgctg	3300
	gtgaaagtaa	aagatgctga	agatcagttg	ggtgcacgag	tgggttacat	cgaactggat	3360
30	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	aacgttttcc	aatgatgagc	3420
	acttttaaa	ttctgctatg	tggcgcggtg	ttatcccgtg	ttgacgcggg	gcaagagcaa	3480
	ctcggtcgcc	gcatacacta	ttctcagaat	gaacttggtg	agtactcacc	agtcacagaa	3540
	aagcatctta	cggatggcat	gacagtaaga	gaattatgca	gtgctgccat	aaccatgagt	3600
	gataacactg	cggccaactt	acttctgaca	acgatcggag	gaccgaagga	gctaaccgct	3660
35	tttttgcaca	acatggggga	tcagttaact	cgcttgatc	gttgggaacc	ggagctgaat	3720
	gaagccatac	caaacgacga	gcgtgacacc	acgatgcctg	tagcaatggc	aacaacggtg	3780
	cgcaaactat	taactggoga	actacttact	ctagcttccc	ggcaacaatt	aatagactgg	3840
	atggaggcgg	ataaagttag	aggaccactt	ctgcgctcgg	cccttccggc	tggctgggtt	3900
	attgctgata	aaatctggagc	cggtgagcgt	gggtctcgcg	gtatcattgc	agcactgggg	3960
40	ccagatggta	agccctcccg	tatcgtagtt	atctacacga	cggggagtca	ggcaactatg	4020
	gatgaacgaa	atagacagat	cgctgagata	ggtgcctcac	tgattaaagca	ttggtaactg	4080
	tcagaccaag	tttactcata	tatacttttag	attgatttaa	aacttcattt	tttaatttaaa	4140
	aggatctagg	tgaagatcct	ttttgataat	ctcatgacca	aaatccctta	acgtgagttt	4200
45	tcgttccact	cccgtagaaa	cccgtagaaa	aagatcaaa	gatcttcttg	agatcccttt	4260
	tttctgcgcg	taatctgctg	cttgcaaaaca	aaaaaaccac	cgctaccagc	ggtgggtttgt	4320
	ttgccggatc	aagagctacc	aactcttttt	ccgaaggtaa	ctggcttcag	cagagcgag	4380
	ataccaataa	ctgtccctct	agtgtagccg	tagttaggcc	accacttcaa	gaactctgta	4440
	gcaccgccta	catacctcgc	tctgtctaac	ctgttaaccag	tggctgctgc	cagtggcgat	4500
	aagtcgtgtc	ttaccgggtt	ggactcaaga	cgatagttac	cggataaggc	gcagcggtcg	4560
50	ggctgaacgg	ggggttcgtg	cacacagccc	agcttgggagc	gaacgaccta	caccgaactg	4620
	agatacctac	agcgtgagct	atgagaaagc	gccacgcttc	ccgaaggggag	aaaggcggac	4680
	aggtatccgg	taagcggcag	ggtcggaaca	ggagagcgca	cgaggagct	tccaggggga	4740
	aacgcctggt	atctttatag	tctgtcgggg	tttcgccacc	tctgacttga	gcgtcgattt	4800
	ttgtgatgct	cgtcaggggg	gcggagccta	tggaaaaacg	ccagcaacgc	ggccttttta	4860
55	cggttctctg	ccttttctgt	gccttttctg	cacatgttct	ttcctgcgtt	atccccgat	4920
	tctgtggata	accgtattac	cgcttttgag	tgagctgata	cgctcgcgcg	cagccgaacg	4980
	accgagcgca	gcgagtcagt	gagcgaggaa	gcggaagagc	gccaataacg	caaaaccgct	5040
	ctccccgcgc	gttggccgat	tcattaatgc	agctggcacg	acaggtttcc	cgactggaaa	5100
	gcgggcagtg	agcgcaacgc	aattaatgtg	agttagctca	ctcattaggc	accccaggct	5160
60	ttacacttta	tgcttcgggc	togtatgttg	tgtggaattg	tgagcggata	acaatttcac	5220
	acaggaaaca	gctatgacca	tgattacgcc	aagctagccc	gggctagctt	gcctgcctgc	5280
	aggtttaaac						5290

<210> 77

65 <211> 5309

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pCMVXisANNLS

5

<400> 77

	aaacagtcgg	atgtacgggc	cagatatacgg	cgttgacatt	gattattgac	tagttattaa	60
	tagtaatcaa	ttacgggggc	attagttcat	agcccatata	tggagttccg	cgttacataa	120
10	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccgccatt	gacgtcaata	180
	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	atgggtggac	240
	tatttacggt	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	aagtacgccc	300
	cctattgacg	tcaatgacgg	taaatggccc	gcttggcatt	atgcccagta	catgacctta	360
	tgggactttc	ctacttggca	gtacatctac	gtattagtc	tcgctattac	catggtgatg	420
	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	atttccaagt	480
15	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	540
	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	acgggtgggag	600
	gtctatataa	gcagagctct	ctggctaact	agagaaccca	ctgcttactg	gcttatcgaa	660
	attaatacga	ctcactatag	ggagacccaa	gctgactcta	gacttaatta	agcgtttggg	720
	tgagtactcc	ctctcaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	780
20	cgaggaggat	ttgatattca	cctggcccg	ggtgatgcct	ttgaggggtg	ccgcgtccat	840
	ctggtcagaa	aagacaatct	ttttgttgct	aagcttgagg	tgtggcaggc	ttgagatctg	900
	gccatacact	tgagtgaac	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	960
	cagggcggcc	gcaccatgcc	caagaagaag	aggaaggtgc	aaaatcaggg	tcaagacaaa	1020
	tatcaacaag	cotttgcaga	cttagagcca	ctttcatcta	ccgacggcag	ttttctcgcc	1080
25	tcaagtctgc	aagcacagca	gcaagagaaa	cacatgagaa	caaaagtact	acaagacct	1140
	gacaaggtaa	atctgcgttt	gaagtctgca	aagacgaaag	tctcagttcg	agaatctaac	1200
	ggaagtctgc	aattacgagc	aacgttacca	attaaacctg	gagataagga	caccaacggt	1260
	acaggcagaa	agcaatacaa	tctcagcttg	aatatccctg	caaacttggg	tggactgaag	1320
	acggctgagg	aagaagctta	tgaattaggt	aaattaatcg	ctcggaaaac	ctttgaatgg	1380
30	aatgataaat	atttaggcaa	agaagccact	aaaaaagatt	cacaaacaat	aggtgattta	1440
	ctagaaaaat	ttgcagaaga	gtatttttaa	acccataaac	gcaccactaa	aagcgaacat	1500
	accttttttt	actatitttt	ccgcacccaa	cgatatacca	attccaaaga	tttagcaacg	1560
	gcggaaaatc	tcatacaatt	aattgagcaa	atcgataaag	aatgggcgag	atataatgcc	1620
	gccagagcca	tatcagcttt	ttgcataaca	ttcaatatag	aaattgattt	gtcccagtag	1680
35	tcocaaaatg	ctgatcgcaa	ttcgcgcaac	atccccacag	atgcagaaat	actatcagga	1740
	attaccaa	ttgaagacta	tctagttacc	agaggaaatc	aagttaatga	agatgtaaaa	1800
	gatagctggc	aactttggcg	ctggacatat	ggaatgtag	cagtttttgg	tttacgcgcc	1860
	agggaaat	ttattaaccc	taatatgtat	tgggtggtta	gcaaagagaa	tatagacctc	1920
	acatggaaag	tagacaaaga	atgtaaaact	ggtgaaagac	aagcattacc	cttacataaa	1980
40	gaatggattg	atgagtttga	tttaagaaat	ccgaaatatt	tagaaatgct	ggcaacagca	2040
	attagtaaaa	aagataaaac	aaatcatgct	gaaataacag	ccttaactca	cgctattagt	2100
	tgggtggttc	ggaaagtcca	attagatttt	aaacctatg	atttacgtca	cgccgtggga	2160
	atcagagcgc	atatttttag	cataccaatc	aaagcggcgg	ctgataattt	ggggcatagt	2220
	atgcaggttc	atacacaac	ctatcagcgc	tgggtctcgc	tagatatgcg	gaagttagcg	2280
45	attaatcagg	ctttgactaa	gaggaatgaa	tttgaggtga	ttagggagga	gaatgctaaa	2340
	ttgcagatag	aaaatgaaag	gttgaggatg	gaaattgaga	agttaaagat	ggaaatagct	2400
	tataagaata	gttgagcggc	cgcgtcgacc	tcgagatcca	ggcgcggtac	aataaaagat	2460
	cattattttc	aatagatctg	tgtgttggtt	ttttgtgtgc	cttggggggg	ggggaggcca	2520
	gaatgaggcg	cggccaaggg	ggagggggg	gccagaatga	ccttggggga	gggggaggcc	2580
50	agaatgacct	tgggggaggg	ggaggccaga	atgaggcgcg	ccccgggta	ccgagctoga	2640
	attcactggc	cgtcgtttta	caacgtcgtg	actgggaaaa	ccctggcggt	acccaactta	2700
	atcgcccttc	agcacatccc	cctttcgcca	gctggcgtaa	tagogaagag	gcccgcaccg	2760
	atcgcccttc	ccaacagttg	cgagcctga	atggcgaaatg	gcgcctgatg	cggatttttc	2820
	tccttacgca	tctgtgcggt	atttcacacc	gcatatggtg	cactctcagt	acaatctgct	2880
55	ctgatgcgc	atagttaaag	cagccccgac	acccgccaac	acccgctgac	gcgcccgtgac	2940
	gggcttgtct	gctccgggca	tcgctttaca	gacaagctgt	gaccgtctcc	gggagctgca	3000
	tgtgtcagag	gttttcaccg	tcatacccca	aacgcgcgag	acgaaagggc	ctcgtgatac	3060
	gcctattttt	ataggttaat	gtcatgataa	taatggtttc	ttagacgtca	ggtggcaact	3120
	ttcggggaaa	tgtgcgcgga	acccctattt	gtttattttt	ctaaatacat	tcaaatatgt	3180
60	atccgctcat	gagacaataa	ccctgataaa	tgcctcaata	atattgaaaa	aggaagagta	3240
	tgagtattca	acatttccgt	gtcgccctta	ttcccttttt	tgcggcattt	tgccttccctg	3300
	ttttgtctca	cccagaaacg	ctgggtgaaa	taaaagatgc	tgaagatcag	ttgggtgcac	3360
	gagtgggtta	cctogaactg	gatctcaaca	gcggttaagat	ccttgagagt	tttcgccccg	3420
	aagaacgttt	tccaattgat	agcactttta	aagttctgct	atgtggcgcg	gtattatccc	3480
65	gtattgacgc	cgggcaagag	caactcggtc	gcgcataca	ctattctcag	aatgacttgg	3540
	ttgagtactc	accagtcaca	gaaaagcatc	ttacggatgg	catgacagta	agagaattat	3600
	gcagtgtctg	cataaccatg	agtataaca	ctgcggccaa	cttacttctg	acaacgatcg	3660

	gaggaccgaa	ggagctaacc	gctttttttgc	acaacatggg	ggatcatgta	actcgccttg	3720
	atcgttggga	accggagctg	aatgaagcca	taccaaaccg	cgagcgtgac	accacgatgc	3780
	ctgtagcaat	ggcaacaacg	ttgcgcaaac	tattaaactgg	cgaactactt	actctagctt	3840
5	cccggaaca	attaatagac	tggatggagg	cggataaagt	tgcaggacca	cttctgcgct	3900
	cggcccttcc	ggctggctgg	tttattgctg	ataaatctgg	agccgggtgag	cgtgggtctc	3960
	gcggatatcat	tgcagcactg	gggccagatg	gtaagccctc	ccgtatcgta	gttatctaca	4020
	cgacggggag	tcaggcaact	atggatgaac	gaaatagaca	gatcgctgag	atagggtgct	4080
	cactgattaa	gcattggtaa	ctgtcagacc	aagtttactc	atatatactt	tagattgatt	4140
10	taaaaacttca	tttttaattt	aaaaggatct	aggtgaagat	cctttttgat	aatctcatga	4200
	ccaaaatccc	ttaacgtgag	ttttcgttcc	actgagcgtc	agaccccgta	gaaaagatca	4260
	aaggatcttc	ttgagatcct	ttttttctgc	gcgtaatctg	ctgcttgcaa	acaaaaaac	4320
	caccgctacc	agcgggtggt	tgtttgccgg	atcaagagct	accaactctt	tttcogaagg	4380
	taactggctt	cagcagagcg	cagataccaa	atactgtcct	tctagtgtag	ccgtagttag	4440
15	gccaccactt	caagaactct	gtagcaccgc	ctacatacct	cgctctgcta	atcctgttac	4500
	cagtggctgc	tgccagtggc	gataagtcgt	gtcttaccgg	gttgactca	agacgatagt	4560
	taccggataa	ggcgcagcgg	tcgggctgaa	cggggggttc	gtgcacacag	cccagcttgg	4620
	agcgaacgac	ctacaccgaa	ctgagatacc	tacagcgtga	gctatgagaa	agcgccacgc	4680
	ttcccgaagg	gagaaaggcg	gacaggatc	cggtaagcgg	cagggtcgga	acaggagagc	4740
20	gcacgagggg	gcttccaggg	ggaaacgcct	ggatatctta	tagtcctgtc	gggtttcgcc	4800
	acctctgact	tgagcgtcga	tttttgtgat	gctcgtcagg	ggggcgagc	ctatggaaaa	4860
	acgccagcaa	cgcggccttt	ttacggttcc	tgcccttttg	ctggcctttt	gctcacatgt	4920
	tctttcctgc	gttatccctt	gattctgtgg	ataaccgtat	taccgccttt	gagtgaagctg	4980
	ataccgctcg	ccgcagccga	acgaccgagc	gcagcgagtc	agtgaagcag	gaagcggag	5040
25	agcgcccaat	acgcaaacgg	cctctccccc	cgcgttggcc	gattcattaa	tgagctggc	5100
	acgacaggtt	tcccgactgg	aaagcgggca	gtgagcgcaa	cgcaattaat	gtgagttagc	5160
	tactcatta	ggcaccocag	gotttacaat	ttatgcttcc	ggctcgtatg	ttgtgtggaa	5220
	ttgtgagcgg	ataacaattt	cacacaggaa	acagctatga	ccatgattac	gccaagctag	5280
	cccgggctag	cttgcatgcc	tgcagggtt				5309
30	<210> 78						
	<211> 7608						
	<212> DNA						
	<213> Artificial Sequence						
35	<220>						
	<223> Description of Artificial Sequence: vector						
	pPGKnifD						
40	<400> 78						
	tcgaggaatt	ctaccgggta	ggggaggcgc	ttttcccaag	gcagtctgga	gcatgcgctt	60
	tagcagcccc	gctgggcaact	tgccgctaca	caagtggcct	ctggctcgca	cacattccac	120
	atccaccggg	aggcgccaac	cggtccggtt	ctttgggtggc	cccttcgcgc	cacctctac	180
45	tcctccccta	gtcaggaagt	tccccccgc	cccgagctc	gcgtcgtgca	ggacgtgaca	240
	aatggaagta	gcacgtctca	ctagtctcgt	gcagatggac	agcaccgctg	agcaatggaa	300
	gcgggtaggg	ctttggggca	gcggccaata	gcagctttgc	tccttcgctt	tctgggtcca	360
	gaggctggga	aggggtgggt	cggggggcgg	gctcaggggc	gggctcaggg	gcggggcggg	420
	cgcccgaagg	tcctccggag	gcccggcatt	ctgcacgctt	caaaagcgca	cgtctgccgc	480
50	gctgtttctc	tcttctcctc	ctccgggcct	ttcgacctgc	agcccggtag	agttcgaatg	540
	gctcttccct	tccgtcaaat	gcactcttgg	gattactccg	aacctagcga	tggggtgcaa	600
	atgtcagatc	agataagttc	gaataacttc	gtatagcata	cattatacga	agttataagc	660
	ttgcatgcct	gcaggtcggc	cgccacgacc	ggccggcccg	tgccgccacc	atcccttgac	720
	ccacgcctct	gacctctcac	aaggagacga	ccttccatga	ccgagtacaa	gccacgggtg	780
55	cgcttcgcca	cccgcgacga	cgtcccccg	gccgtacgca	ccctcgccgc	cggttcgcc	840
	gactaccccg	ccacgcgcca	caccgtcgac	ccggaccgcc	acatcgagcg	ggtcaccgag	900
	ctgcaagaac	tcttctctac	gcgcgtcggg	ctcgacatcg	gcaagggtgtg	ggtcgcggac	960
	gacggcgccg	cgggtggcgg	ctggaccacg	ccggagagcg	tcgaagcggg	ggcgggtgtc	1020
	gccgagatcg	gcccgcgcac	ggccaggttg	agcggttccc	ggctggccgc	gcagcaacgc	1080
60	atggaaggcc	tcctggcgcc	gcaccggccc	aaggagcccg	cgtggttcc	ggccaccgtc	1140
	ggcgtctcgc	ccgaccacca	gggcaagggt	ctgggcagcg	ccgtcgtgct	ccccggagtg	1200
	gaggcggccg	agcgcgcggg	ggtgcccgc	ttcttgagga	cctccgcgcc	ccgcaacctc	1260
	cccttctacg	agcggctcgg	cttaccgctc	accgccgacg	tcgagtgcgc	gaaggaccgc	1320
	gcgacctggt	gcattgacccg	caagcccggg	gcctgacgcc	cgccccacga	ccgcagcgcg	1380
65	ccgaccgaaa	ggagcgcaag	acccccatgg	tcgcagcgaa	gccgaccggg	gcggcccgcc	1440
	cgaccccgca	ccgcgccccg	agggccaccc	actctagagg	atcataatca	gccataccac	1500
	atgtgtagag	gttttaacttg	ctttaaaaaa	cctcccacac	ctccccctga	acctgaaaca	1560
	taaaatgaat	gcaattgttg	ttgttaactt	gtttatttga	gcttataatg	gttacaataa	1620

	aagcaatagc	atcacaaatt	tcacaaataa	agcatttttt	tcactgcatt	ctagttgtgg	1680
	tttgtccaaa	ctcatcaatg	tatcttatca	tgtctggatc	cagctgttga	aagctattaa	1740
	accacaaaaa	ggattactcc	ggcccttatc	acggttacga	cggattttga	tccataactt	1800
5	cgatatagcat	acattatacg	aagttatacc	gggccaccat	ggtcgcgagt	agcttggcac	1860
	tggccgtcgt	tttacaacgt	cgtgactggg	aaaaccctgg	cgttacccaa	cttaatcgcc	1920
	ttgcagcaca	tcccccttcc	gocagctggc	gtaatagoga	agaggcccg	accgatcgcc	1980
	cttcccaaca	gttgogcagc	ctgaatggcg	aatggcgctt	tgcctgggtt	ccggcaccag	2040
	aagcgggtgcc	ggaaagctgg	ctggagtgcg	atcttctctga	ggccgatact	gtcgtcgtcc	2100
10	cctcaaaactg	gcagatgcac	ggttacgatg	cgcccatcta	caccaacgta	acctatccca	2160
	ttacgggtcaa	tccgcgcgtt	gttcccacgg	agaatccgac	gggttggttac	tcgctcacat	2220
	ttaatgttga	tgaagctggg	ctacaggaag	gccagacgcg	aattattttt	gatggcggtta	2280
	actcggcggtt	tcactctgtg	tgcaacgggg	gctgggtcgg	ttacggccag	gacagctcgtt	2340
	tgcgctctga	atttgaocctg	agcgcatttt	tacgycgcgg	agaaaaaccg	ctcgcggtga	2400
15	tgggtgctgcg	ttggagtgcg	ggcagttatc	tggaagatca	ggatatgtgg	cggatgagcg	2460
	gcattttcccg	tgacgtctcg	ttgctgcata	aaccgactac	acaaatcagc	gattttccatg	2520
	ttgccactcg	ctttaatgat	gatttcagcc	gcgctgtact	ggaggctgaa	gttcagatgt	2580
	gcggcgagtt	gcgtgactac	ctacgggttaa	cagtttcttt	atggcagggg	gaaacgcagg	2640
	tcgccagcgg	gcaccgcgct	ttcggcggtg	aaattctcga	tgagcgtggg	ggttatcccg	2700
	atcgcgtcac	actacgctcg	aacgtgaaa	accgaaaact	gtggagcgcc	gaaatccga	2760
20	atctctatcg	tgcggtggtt	gaactgcaca	ccgcgcgacg	cacgctgatt	gaagcagaag	2820
	cctgcgatgt	cggtttccgc	gaggtgcgga	ttgaaaatgg	tctgctgctg	ctgaacggca	2880
	agccgttgct	gattcagagg	gttaaccgtc	acgagcatca	tcctctgcat	ggtcaggtca	2940
	tggatgagca	gacgatgggt	caggatatcc	tgctgatgaa	gcagaacaac	tttaacgccg	3000
	tgcgctgttc	gcattatccg	aaccatccgc	tgtggtacac	gctgtgcgac	cgttacggcc	3060
25	tgtatgtggt	ggatgaagcc	aatattgaaa	ccacgycat	ggtgccaatg	aatcgtctga	3120
	ccgatgatcc	gcgctggcta	ccggcgatga	gcgaacgcgt	aacgcgaatg	gtgcagcgcg	3180
	atcgtaatca	cccgagtgtg	atcatctggt	cgctggggaa	tgaatcaggc	cacggcgcta	3240
	atcacgacgc	gctgtatcgc	tggatcaaat	ctgtcgatcc	ttcccgcccg	gtgcagtatg	3300
	aaggcggcgg	agccgcaccc	acggccaccg	atattatttt	cccgatgtac	gcgcgcgtgg	3360
30	atgaagacca	gcccttcccg	gctgtgccga	aattggtccat	caaaaaatgg	ctttcgctac	3420
	ctggagagac	gcgcccgcgt	atcctttgcg	aatacgccca	cgcgatgggt	aacagtcctg	3480
	gcggtttcgc	taaatactgg	caggcggttc	gtcagtatcc	ccgtttacag	ggcggcttcg	3540
	tctgggactg	ggtggatcag	tcgctgatta	aatatgatga	aaacggcaac	ccgtggctcg	3600
35	cttacggcgg	tgattttggc	gatacgccga	acgatcgcca	gttctgtatg	aacggtctgg	3660
	tctttgccga	ccgcacgcg	catccagcgc	tgacggaagc	aaaacaccag	cagcagtttt	3720
	tccagttccg	tttatccggg	caaaccatcg	aagtgcaccg	cgaatacctg	ttccgtcata	3780
	gcgataacga	gctcctgcac	tggatggtgg	cgctggatgg	taagccgctg	gcaagcgggtg	3840
	aagtgcctct	ggatgtcgc	ccacaaggta	aacagttgat	tgaactgcct	gaactaccgc	3900
	agccggagag	cgccgggcaa	ctctgggtca	cagtaacgcg	agtgcacccg	aacgcgaccg	3960
40	catggtcaga	agccggggc	atcagcgcct	ggcagcagtg	gcgtctggcg	gaaaacctca	4020
	gtgtgacgct	ccccgcgcg	tcccacggca	tccgcactct	gaccaccagc	gaaatggatt	4080
	tttgcatcga	gctgggtaat	aagcgttggc	aatttaaccg	ccagtcaggc	tttctttcac	4140
	agatgtggat	tggcgataaa	aaacaactgc	tgacgcccgt	gcgcgatcag	ttcaccctgt	4200
45	caccgctgga	taacgacatt	ggcgtaagt	aagcgaccgg	cattgaccct	aacgcctggg	4260
	tcgaacgctg	gaaggcggcg	ggccattacc	aggccgaagc	agcgttgttg	cagtgcacgg	4320
	cgataacact	tgctgatgcg	gtgctgatta	cgaccgctca	cgcgtggcag	catcagggga	4380
	aaaccttatt	tatcagccgg	aaaacctacc	ggattgatgg	tagtgggtcaa	atggcgatta	4440
	ccgttgatgt	tgaagtggcg	agcgatacac	gcgatccggc	gcggattggc	ctgaactgcc	4500
50	agctggcgca	ggtagcagag	cgggtaaaact	ggctcggatt	agggccgcaa	gaaaactatc	4560
	ccgaccgcct	tactgcgcgc	tgttttgacc	gctgggatct	gccattgtca	gacatgtata	4620
	ccccgtacgt	cttcccgagc	gaaaaacggt	tgcgctgcgg	gacgcgcgaa	ttgaattatg	4680
	gcccacacca	gtggcgcgcc	gacttccagt	tcaacatcag	ccgctacagt	caacagcaac	4740
	tgatggaaac	cagccatcgc	catctgctgc	acgcggaaga	aggcacatgg	ctgaatatcg	4800
55	acggtttcca	tatggggatt	ggtggcgacg	actcctggag	cccgtcagta	tcggcggaat	4860
	tccagctgag	cgccggctgc	taccattacc	agttggtctg	gtgtcaaaaa	taataataac	4920
	cgggcagggg	ggatctttgt	gaaggaaact	tacttctgtg	gtgtgacata	attggacaaa	4980
	ctacctacag	agatttaaa	ctctaaggta	aatataaaat	ttttaagtgt	ataatgtgtt	5040
	aaactactga	ttctaattgt	ttgtgtatit	tagattccaa	cctatggaac	tgtgaattgg	5100
60	gagcagtggt	ggaatgccag	atccagacat	gataagatag	attgatgagt	ttggacaaac	5160
	cacaactaga	atgcagtga	aaaaatgctt	tatttgtgaa	atttgtgatg	ctattgtottt	5220
	atttgaacc	attataagct	gcaataaaca	agttaacaac	aacaattgca	ttcattttat	5280
	gtttcaggtt	cagggggagg	tgtgggaggt	tttttaaaag	aagtaaaacc	tctacaaatg	5340
	tggatggct	gattatgatc	tgcggcccgca	gggcctcgtg	atacgcctat	ttttataggt	5400
	taatgtcatg	ataataatg	ttcttatagc	ctcaggtggc	acttttcggg	gaaatgtgcg	5460
65	cggaaacccct	atttgtttt	ttttctaaat	acattcaaat	atgtatccgc	tcattgagaca	5520
	ataaccctga	taaagtcttc	aataatattg	aaaaaggaag	agtatgagta	ttcaacattt	5580
	ccgtgtcgcc	cttattccct	tttttgcggc	attttgcctt	cctgtttttg	ctcaccacaga	5640

	aacgctgggtg	aaagtaaaag	atgctgaaga	tcagttgggt	gcacgagtg	gttacatcga	5700
	actggatctc	aacagcggta	agatccttga	gagttttcgc	cccgaagaac	gttttccaat	5760
	gatgagcaact	tttaaaagttc	tgctatgtgg	cgcggtatta	tcccgatttg	acgcggggca	5820
5	agagcaactc	ggtcgccgca	tacactattc	tcagaatgac	ttggttgagt	actcaccagt	5880
	cacagaaaag	catcttacgg	atggcatgac	agtaagagaa	ttatgcagt	ctgccataac	5940
	catgagtgat	aaacttgccg	ccaacttact	tctgacaacg	atcggaggac	cgaaggagct	6000
	aacgcgtttt	ttgcacaaca	tgggggatca	tgtaactcgc	cttgatcggt	gggaaccgga	6060
	gctgaatgaa	gccataccaa	acgacgagcg	tgacaccacg	atgcctgtag	caatggcaac	6120
10	aacgttgccg	aaactattaa	ctggcgaaact	acttactcta	gcttcccggc	aacaattaat	6180
	agactggatg	gaggcggata	aaagtgcagg	accacttctg	cgctcggccc	ttccggctgg	6240
	ctggtttatt	gctgataaat	ctggagccgg	tgagcgtggg	tctcgcggta	tcattgcagc	6300
	actggggcca	gatggttaag	cctcccgat	cgtagttatc	tacaogacgg	ggagctcagc	6360
	aactatggat	gaacgaata	gacagatcgc	tgagataggt	gcctcaactga	ttaagcattg	6420
	gtaactgtca	gaccaagttt	actcatatat	acttttagatt	gatitaaac	ttcattttta	6480
15	atttaaaagg	atctaggtga	agatcctttt	tgataatctc	atgacaaaa	tcccttaacg	6540
	tgagttttcg	ttccactgag	cgtcagaccc	cgtagaaaaag	atcaaaggat	cttcttgaga	6600
	tccttttttt	ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccacgcg	taccagcgg	6660
	ggtttggttg	coggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	6720
	agcgcagata	ccaaatactg	tccttctagt	gtagccgtag	ttaggccacc	acttcaagaa	6780
20	ctctgtagca	ccgcctacat	acctcgctct	gctaattcctg	ttaccagtgg	ctgctgccag	6840
	tggcgataag	tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	6900
	gcggtcgggc	tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	6960
	cgaactgaga	tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aagggagaaa	7020
	ggcggacagg	tatccggtaa	gcggcagggt	cggacagga	gagcgacga	gggagcttcc	7080
25	agggggaaac	gcctggtatc	tttatagttc	tgtcgggttt	cgccacctct	gacttgagcg	7140
	tcgatttttg	tgatgctcgt	caggggggcg	gagcctatgg	aaaaacgcca	gcaacgcggc	7200
	cttttttacgg	ttcctggcct	tttgcctggc	ttttgctcac	atgttctttc	ctgcgttctc	7260
	ccctgattct	gtggataacc	gtattaccgc	ctttgagtga	gotgataccg	ctcgcgcgag	7320
	ccgaacgacc	gagcgcagcg	agtcagtgag	cgaggagcgc	gaagagcgcc	caatacgcga	7380
30	accgcctctc	cccgcgcgtt	ggcggattca	ttaatgcagc	tggaacgaca	ggtttcccgga	7440
	ctggaaagcg	ggcagtgagc	gcaacgcaat	taatgtgagt	tagctcactc	attaggcacc	7500
	ccaggcttta	cactttatgc	ttccggctcg	tatgttgtgt	ggaattgtga	gcggataaca	7560
	atttcacaca	ggaaacagct	atgaccatga	ttacgccaag	ctggcgcg		7608
35	<210> 79						
	<211> 7523						
	<212> DNA						
	<213> Artificial Sequence						
40	<220>						
	<223> Description of Artificial Sequence: vector						
	pPGKnifD3'						
45	<400> 79						
	tcgagggaatt	ctaccgggta	ggggaggcgc	ttttcccaag	gcagtctgga	gcatgcgctt	60
	tagcagcccc	gctgggcact	tggcgctaca	caagtggcct	ctggctcgca	cacattccac	120
	atccaccgggt	aggcgccaac	cggctccgtt	ctttgggtgg	cccttcgcgc	caccttctac	180
50	tcctccccta	gtcaggaagt	tccccccgc	ccgcagctc	gcgtcgtgca	ggacgtgaca	240
	aatggaagta	gcacgtctca	ctagtctcgt	gcagatggac	agcaccgctg	agcaatggaa	300
	gcggttaggc	ctttggggca	ggggccaata	gcagctttgc	tccttcgctt	tctgggctca	360
	gaggctggga	agggtgggt	ccggggcgcg	gctcaggggc	gggctcaggg	gcggggcggg	420
	cgcccgaagg	tcctccggag	gcccggcatt	ctgcacgctt	caaaagcgca	cgtctgccgc	480
	gctgtttctc	tccttcctcat	ctccgggcct	ttcgacctgc	agcccgttac	agttcgaata	540
55	acttcgtata	gcatacatta	tacgaagtta	taagcttgca	tgcttgcaag	tcggccgcca	600
	cgaccggccg	gcccgtgccc	ccaccatccc	ctgacccacg	cccttgaccc	ctcacaagga	660
	gacgaccttc	catgaccgag	tacaagcccc	cggctgcgct	cgccacccgc	gacgacgtcc	720
	cccggccggt	acgcaccctc	gcccgcgctg	tcgcgcacta	ccccgccacg	cgccacacgc	780
	tcgaccggga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	840
60	tcgggctcga	catcggcaag	gtgtgggtcg	cggacgacgg	cgccgcgggtg	gcggtctgga	900
	ccaacgcggga	gagcgtcgaa	gcggggggcg	tgctcgccga	gatcggcccg	cgcatggccg	960
	agttgagcgg	ttcccggctg	gcgcgcgagc	aacagatgga	aggcctcctg	gcgcgcgacc	1020
	ggcccaagga	gcccgcgtgg	ttcctggcca	ccgtcggcgt	ctcgcgcgac	caccagggga	1080
	aggtctctgg	cagcgcgctc	gtcctgcccc	gagtgaggcg	ggccgagcgc	gccggggtgc	1140
65	ccgccttctc	ggagacctcc	gcgcgccgca	acctcccttt	ctacgagcgg	ctcggcttca	1200
	ccgtcacccg	cgacgtcgag	tgcccgaagg	accgcgcgac	ctggtgcatg	accgcgaagc	1260
	ccggtgcctg	acgcccgcgc	cacgacccgc	agcgcgccgac	cgaaaggagc	gcacgacccc	1320

	atgggtccga	ccgaagccga	ccggggcgcc	ccgcgcgacc	ccgcacccgc	ccccgaggcc	1380
	caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1440
	aaaaacotcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tgttggtgtt	1500
5	aacttggtta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1560
	aataaagcat	ttttttcact	gcattctagt	tgtggtttgt	ccaaactcat	caatgtatct	1620
	tatcatgtct	ggatccagct	gttgaagct	atthaaccac	aaaaaggatt	actccggccc	1680
	ttatcacggt	tacgacggat	ttggatccat	aacttctgtat	agcatacatt	atacgaagtt	1740
	ataccggggc	accatggtcg	cgagtagctt	ggcaactggc	gtcgttttac	aacgtcgtga	1800
10	ctgggaaaac	cctggcggtta	cccaacttaa	tgcctttgca	gcacatcccc	ctttcgccag	1860
	ctggcgtaat	agcgaagagg	cccgacccga	tgcctcttcc	caacagttgc	gcagcctgaa	1920
	tggcgaatgg	cgctttgcct	ggtttccggc	accagaagcg	gtgcccggaa	gctggctgga	1980
	gtgcgatctt	cctgaggccg	atactgtcgt	cgtccctcca	aactggcaga	tgcacgggtta	2040
	cgatgcgccc	atctacacca	acgtaacctt	tcccattacg	gtcaatccgc	cgtttcttcc	2100
	cacggagaat	ccgacggggt	gttactcgct	cacattttaat	gttgatgaaa	gctggctaca	2160
15	ggaaggccag	acgcgaatta	tttttgatgg	cgtttaactcg	gcgtttcatc	tgtggtgcaa	2220
	cgggcgctgg	gtcggttacg	gccaggacag	tcggtttgocg	tctgaatttg	acctgagcgc	2280
	attttttacgc	gccggagaaa	accgcctcgc	ggtgatggtg	ctgcgttgga	gtgacggcag	2340
	ttatctggaa	gatcagata	tgtggcggtt	gagcggcatt	ttcctgtgacg	tctctgtgct	2400
	gcataaacgg	actacacaaa	tcagcgattt	ccatggttgc	actcgtttta	atgatgattt	2460
20	cagccgcgct	gtactggagg	ctgaagtcca	gatgtgcggc	gagttgcgtg	actacctacg	2520
	ggtaacagtt	tctttatggc	aggggtgaac	gcagggtcgcc	agcggcacccg	cgcttttcgg	2580
	cggtgaaatt	atcgatgagc	gtggtggtta	tgcgatcgcc	gtcacactac	gtctgaacgt	2640
	cgaaaacccg	aaactgtgga	gcgcggaaat	ccogaatctc	tatcgtgcgg	tgggtgaact	2700
	gcacaccgcc	gacggcacgc	tgattgaagc	agaagcctgc	gatgtcgggt	tccgcgaggt	2760
25	gcggattgaa	aatggctctgc	tgtctgtgaa	cggaagccg	ttgctgattc	gaggcgttaa	2820
	ccgtcacgag	catcatcctc	tgcattggtca	ggatcatggat	gagcagacga	tgggtgcagga	2880
	tatcctgctg	atgaagcaga	acaactttta	cgccgtgcgc	tgttcgcatt	atccgaacca	2940
	tccgtctgtg	tacacgctgt	gcgaccgcta	cgccctgtat	gtggtggatg	aagccaatat	3000
	tgaacccac	ggcatggtgc	caatgaatcg	tctgaccgat	gatccgcgct	ggctaccggc	3060
30	gatgagcgaa	cgcgtaacgc	gaatggtgca	gcgogatcgt	aatcacccga	gtgtgatcat	3120
	ctgggtcgctg	gggaatgaat	caggccacgg	cgctaatac	gacgcgctgt	atcgcctggat	3180
	caaactctgtc	gatccttccc	gcccggtgca	gtatgaaggc	ggcggagccg	acaccacggc	3240
	caccgatatt	atttgcccga	tgtacgcgcg	cgtggatgaa	gaccagccct	tcccggctgt	3300
	gccgaaatgg	tccatcaaaa	aatggctttc	gctacctgga	gagacgcgcc	cgctgatect	3360
35	ttgcgaatac	gcccacgcga	tgggttaacg	tcttggcggt	ttcgctaaat	actggcaggc	3420
	gtttcgtcag	tatccccggt	tacaggcgcg	cttctgtctg	gactgggtgg	atcagtcgct	3480
	gattaaatat	gatgaaaacg	gcaaccocgtg	gtcggcttac	ggcgggtgatt	ttggcgatac	3540
	gccgaacgat	cgccagttct	gtatgaacgg	tctggctctt	gccgaccgca	cgccgcattcc	3600
	agcgtcgacg	gaagcaaaac	accagcagca	gtttttccag	ttccgtttat	ccgggcaaac	3660
40	catcgaagtg	accagcgaat	acctgttccg	tcatagcgat	aacgagctcc	tgcactggat	3720
	gggtggcgctg	cggtgtaagc	cgctggcaag	ccgtgaagtg	cctctggatg	tgcgtccaca	3780
	aggtaaacag	ttgattgaac	tgcctgaact	accgcagccg	gagagcgccg	ggcaactctg	3840
	gctcacagta	cgcgtagtgc	aaocgaacgc	gaccgcattg	tcagaagccg	ggcacatcag	3900
45	cgctggcgag	cagtggcgtc	tggcggaaaa	cctcagtggtg	acgctccccg	ccgcgtccca	3960
	cgccatcccg	catctgacca	ccagcgaaat	ggattttttgc	atcgagctgg	gtaataagcg	4020
	ttggcaattt	aacccgcagt	caggctttct	ttcacagatg	tggattggcg	ataaaaaaca	4080
	actgtctagc	ccgctgcgcg	atcagttcac	cogtgcacccg	ctggataacg	acattggcgt	4140
	aagtgaagcg	accgcgattg	accctaacgc	ctgggtcgaa	cgctggaagg	cgccggggcca	4200
	ttaccaggcc	gaagcagcgt	tgttgcaagt	cacggcagat	acacttgctg	atgcggtgct	4260
50	gattacgacc	gctcacgcgt	ggcagcatca	ggggaaaacc	ttattttatca	gccgaaaaac	4320
	ctaccggatt	gatggtagtg	gtcaaatggc	gattaccggt	gatgttgaag	tggcgagcga	4380
	tacaccgcat	ccggcgcgga	ttggcctgaa	ctgccagctg	gcgcaggtag	cagagcgggt	4440
	aaactggctc	ggattagggc	cgcaagaaaa	ctatcccgcg	cgcttactg	ccgcctgttt	4500
	tgaccgctgg	gatctgccat	tgtcagacat	gtataccccg	tacgtcttcc	cgagcgaaaa	4560
55	cggtctgcgc	tgcgggacgc	gogaattgaa	ttatggccca	caccagtggc	gcggcgactt	4620
	ccagttcaac	atcagccgct	acagtoaaca	gcaactgatg	gaaaccagcc	atcgccatct	4680
	gctgcacgcg	gaagaaggca	catggtgaa	tatcgacggt	ttccatatgg	ggattgggtg	4740
	cgacgactcc	tggagcccg	cagtatcgcc	ggaattccag	ctgagcgccg	gtcgctacca	4800
	ttaccagttg	gtctggtgtg	aaaaataata	ataaccggcg	aggggggatc	tttctgaagg	4860
60	aaccttactt	ctgtggtgtg	acataattgg	acaaactacc	tacagagatt	taaagctcta	4920
	aggtaaatat	aaaattttta	agtgtataat	gtgttaaact	actgattcta	attgtttgtg	4980
	tatttttagat	tccaacctat	ggaactgatg	aatgggagca	gtggtggaat	gccagatcca	5040
	gacatgataa	gatacattga	tgagtttgga	caaaccacaa	ctagaatgca	gtgaaaaaaa	5100
	tgctttattt	gtgaaatttg	tgtgcttatt	gctttatttg	taaccattat	aagctgcaat	5160
65	aaacaagtta	acaacaacaa	ttgcattoat	tttatgtttc	aggttcaggg	ggaggtgtgg	5220
	gaggtttttt	aaagcaagta	aaacctctac	aaatgttgta	tggctgatta	tgatctgcgg	5280
	ccgcaggggc	tcggtgatacg	cctattttta	taggttaatg	toatgataat	aatggtttct	5340

	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	tttatttttc	5400
	taaatacatt	caaatatgta	tccgctcatg	agacaataac	cctgataaat	gcttcaataa	5460
	tattgaaaaa	ggaagagtat	gagtattcaa	catttcctgt	tcgcccttat	tccctttttt	5520
5	gcggcatttt	gccttctgt	ttttgctcac	ccagaaacgc	tggtgaaagt	aaaagatgct	5580
	gaagatcagt	tggtgtcacg	agtgggttac	atcgaactgg	atctcaacag	cggtaagatc	5640
	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	agttctgcta	5700
	tgtggcgcg	tattatcccg	tattgacgcc	gggcaagagc	aactcggtcg	ccgcatacac	5760
	tattctcaga	atgacttgg	tgagtactca	ccagtcacag	aaaagcatct	tacggatggc	5820
10	atgacagtaa	gagaattatg	cagtgtctgc	ataaccatga	gtgataaac	tgcggccaac	5880
	ttactttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	caacatgggg	5940
	gatcatgtaa	ctcgcccttg	tcgttgggaa	cgggagctga	atgaagccat	accaaaccgac	6000
	gagcgtgaca	ccacgatgcc	tgtagcaact	ccaacaacgt	tgcgcaaact	attaactggc	6060
	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	ggataaagtt	6120
15	gcaggaccac	ttctgcgctc	ggcccttcgc	gctggctgg	ttattgctga	taaatctgga	6180
	gccggtgagc	gtgggtctcg	cggtatcatt	gcagcactgg	ggccagatgg	taagccctcc	6240
	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggtgaacg	aatagacag	6300
	atcgtcgtga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	agtttactca	6360
	tatatacttt	agattgattt	aaaacttcct	ttttaattta	aaaggatcta	ggtgaagata	6420
20	cttttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	6480
	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgog	cgtaatctgc	6540
	tgcttgcaaa	caaaaaaacc	accgctacca	gcggtgggtt	gtttgcggga	tcaagagcta	6600
	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	6660
	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcacccgc	tacatacctc	6720
25	gctctgctaa	tcctgttacc	agtggtgct	gccagtggcg	ataagtctgt	tcttacccgg	6780
	ttggactcaa	gacgatagtt	accggataag	gcgcagcgg	cgggctgaac	ggggggttcg	6840
	tgacacacgc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	6900
	ctatgagaaa	gcgccacgct	tcccgaaagg	agaaaggcgg	acaggtatcc	ggtaaagcgc	6960
	agggtcggaa	caggagagcg	cacgagggag	cttcacgggg	gaaacgcctg	gtatctttat	7020
	agtcctgtgc	ggtttcgcc	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	7080
30	ggcgagagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	7140
	tggtcctttt	ctcacatgtt	cttctctgcg	ttatcccttg	attctgtgga	taaccgtatt	7200
	accgcctttg	agttagctga	taccgctcgc	cgcagccgaa	cgaccgagcg	cagcagtgca	7260
	gtgagcgagg	aagcgggaaga	gcgcccataa	cgaaaccgc	ctctccccgc	gcgttgcccg	7320
35	attcattaat	gcagctggca	cgacaggttt	cccgactgga	aagcgggcag	tgagcgcaac	7380
	gcaattaatg	tgagttagct	cactcattag	gcaccccgag	ctttacactt	tatgcttccg	7440
	gctcgtatgt	tgtgtggaat	tgtgagcgga	taacaatttc	acacaggaaa	cagctatgac	7500
	catgattacg	ccaagctggc	gcg				7523
40	<210> 80						
	<211> 4754						
	<212> DNA						
	<213> Artificial Sequence						
45	<220>						
	<223> Description of Artificial Sequence: vector						
	pRK55pbsC31-Re						
50	<400> 80						
	ggcgcccgca	tatgacacaa	gggttgtgta	ccgggtgga	cacgtacgog	ggtgcttacg	60
	accgtcagtc	gcgcgagcgc	gagaattcga	gcgcagcaag	cccagcgaca	cagcgtagcg	120
	ccaacgaaga	caaggcgccg	gaccttcagc	gcgaagtcca	gcgcgacggg	ggcgggttca	180
	ggttcgtcgg	gcatttcagc	gaagcgccgg	gcacgtcgcc	gttcgggacg	gcggagcgcc	240
55	cggagttcga	acgcatacctg	aacgaatgcc	gcgcggggcg	gotcaacatg	atcattgtct	300
	atgacgtgtc	gcgcttctcg	cgctgaagg	tcattggacgc	gattccgatt	gtctcggaat	360
	tgotcgccct	ggcggtgacg	attgtttcca	ctcaggaagg	cgtcttcogg	cagggaaaacg	420
	tcattggacct	gattcacctg	attatgcggc	tcgacgcgtc	gcacaaagaa	tcttcgctga	480
	agtcggcgaa	gattctcgac	acgaagaacc	ttcagcgcca	attggcgggg	tacgtcggcg	540
60	ggaaggcgcc	ttacggcttc	gagcttggtt	cggagacgaa	ggagatcacg	cgcaacggcc	600
	gaatgggtcaa	tgtcgtcatc	aacaagcttg	cgcactcgac	cactccccctt	accggaccct	660
	tcgagttcga	gcccagcgta	atccggtggt	ggtggcgtga	gatcaagacg	cacaaacacc	720
	ttcccttcaa	gcggggcagt	caagcgccca	ttcaccgggg	cagcatcacg	gggcttttga	780
	agcgcatgga	cgtgacgcc	gtgccgaccc	ggggcgagac	gattgggaag	aagaccgctt	840
65	caagcgctga	ggaccggcca	accgttatgc	gaatccttcg	ggaccggcgt	attgogggct	900
	tcggcgctga	ggtgatctac	aagaagaagc	cggacggcac	gccgaccacg	aagattgagg	960
	gttaccgcat	tcagcgcgac	ccgatcacgc	tcgggcccgt	cgagcttgat	tgcggaccga	1020
	tcacgcagcc	cgtgagtg	tatgagcttc	aggcgtggtt	ggacggcagg	ggcgcgccga	1080

	aggggctttc	ccgggggcaa	gccattctgt	ccgccatgga	caagctgtac	tgcgagtgtg	1140
	gogccgtcat	gacttcgaag	cgcggggaag	aatogatcaa	ggactcttac	cgctgccgtc	1200
	gccggaaggt	ggtcgacccg	tccgcacctg	ggcagcacga	aggcacgtgc	aacgtcagca	1260
5	tggcggaact	cgacaagttc	gttgcggaac	gcatcttcaa	caagatcagg	cacgccgaag	1320
	gogacgaaga	gacgttggcg	cttctgtggg	aagccgcccc	acgtttcggc	aagctcactg	1380
	aggcgccctga	gaagagcggc	gaacggggcg	accttgttgc	ggagcgcgcc	gacgccctga	1440
	acgcccttga	agagctgtac	gaagaccgcg	cggcaggcgc	gtacgacgga	cccgttggca	1500
	ggaagcactt	ccggaagcaa	caggcagcgc	tgaagctccg	gcagcaaggg	gcggaagagc	1560
	ggcttgccga	acttgaagcc	gccgaagccc	cgaagcttcc	ccttgaccaa	tggttccccg	1620
10	aagacgcccga	cgctgacccg	accggcccta	agtcgtggtg	ggggcgcgcg	tcagttagacg	1680
	acaagcgcggt	gttcgtcggg	ctcttcgtag	acaagatcgt	tgtaacgaag	tcgactacgg	1740
	gcaggggggca	gggaacgccc	atcgagaagc	gcgcttcgat	cacgtgggcg	aagccgcccga	1800
	ccgacgacga	cgaagacgac	gcccaggacg	gcacggaaga	cgtagcgggc	taggcggcgc	1860
	ccgggctcga	ggggggggccc	ggtacccagc	ttttgttccc	tttagtgagg	gttaatttccg	1920
15	agcttggcgt	aatcatggtc	atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	1980
	ccacacaaca	tacgagccgg	aagcataaag	tgtaaagcct	gggggtgccta	atgagttagc	2040
	taactcacat	taattgcgtt	gcgctcactg	cccgctttcc	agtcgggaaa	cctgtcgtgc	2100
	cagctgcatt	aatgaatcgg	ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	2160
	tcogcttcc	cgctcactga	ctcgctgcgc	tcggtcggtc	ggctgcggcg	agcggtatca	2220
20	gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	2280
	atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcgttt	2340
	ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	2400
	cgaaccgccga	caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	2460
	tctcctgttc	cgaccctgcc	gcttacccga	tacotgtccg	cctttctccc	ttcgggaagc	2520
25	gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	2580
	aagctgggct	gtgtgcacga	accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	2640
	tatcgtcttg	agtccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactggg	2700
	aacaggatta	gcagagcgag	gtatgtaggc	gggtgtacag	agttcttgaa	gtggtggcct	2760
	aactacggct	acactagaag	gacagtattt	ggtaattcgc	ctctgctgaa	gccagttacc	2820
30	ttcggaaaaa	gagttggtag	ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcggtggg	2880
	ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	2940
	atotttttcta	cggggctctga	cgctcagttg	aacgaaaact	cacgttaagg	gatttttggtc	3000
	atgagattat	caaaaaggat	cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	3060
	tcaatctaaa	gtatatatga	gtaaaccttg	tctgacagtt	accaatgott	aatcagtgag	3120
35	gcacctatct	cagcgatctg	tctatttctg	tctaccatag	ttgcctgact	ccccgtcgtg	3180
	tagataacta	cgatacggga	gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	3240
	gaaccacgct	caccggctcc	agatttatca	gcaataaacc	agccagccgg	aagggccgag	3300
	cgcagaagtg	gtcctgcaac	tttatccgcc	tccatccagt	ctattaattg	ttgccgggaa	3360
	gctagagtaa	gtagttcgcc	agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	3420
40	atcgtgggtg	cacgctcgtc	gtttggtatg	gcttcaattc	gctccgggtc	ccaacgatca	3480
	aggcgagtta	catgttctgc	aaaaaagcgg	aaaaaagcgg	ttagctcctt	cggtcctccg	3540
	atcgttgtca	gaagtaagtt	ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	3600
	aattctctta	ctgtcatgcc	atccgtaaga	tgctttttctg	tgactggtga	gtactcaacc	3660
	aagtcattct	gagaatagtg	tatgcggcga	cagagttgct	cctgcccggc	gtcaatacgg	3720
45	gataataccg	cgccacatag	cagaacttta	aaagtgtcca	tcattggaaa	acgttcttcg	3780
	gggcgaaaaa	tctcaaggat	cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	3840
	gcaccacaact	gatcttcagc	atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaca	3900
	ggaaggcaaa	atgccgcaaa	aaagggaata	aggcgacac	ggaaatggtg	aatactcata	3960
	ctcttccttt	ttcaatatata	ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	4020
50	atatttgaa	gtatttagaa	aaataaaca	ataggggttc	cgcgcacatt	tccccgaaaa	4080
	gtgccacctta	aattgtaagc	gttaatatatt	tgttaaaaat	cgcgtaaaat	ttttgttaaa	4140
	tcagctcatt	ttttaaccaa	taggcgaaa	tcggcaaaa	cccttataaa	tcaaaagaat	4200
	agaccgagat	agggttgagt	gttgttccag	tttggaacaa	gagtcacta	ttaaagaacg	4260
	tggactccaa	cgtcaaaggg	cgaaaaacgg	tctatcaggg	cgatggccca	ctacgtgaac	4320
55	catcacccta	atcaagtttt	ttggggctga	gggtgccgta	agcactaaat	cggaaacccta	4380
	aaggggagccc	ccgattttaga	gcttgacggg	gaaagccggc	gaacgtggcg	agaaagggaag	4440
	ggaagaaagc	gaaaggagcg	ggcgctaggg	cgctggcaag	tgtagcgggc	acgctgcgcg	4500
	taaccaccac	accgcgcgcg	cttaatgcgc	cgtcacaggg	cgcgtcccat	tcgccattca	4560
	ggctgcccga	ctgttgggaa	ggcgcatcgg	tcggggcctc	ttcgctatta	cgcagctggg	4620
60	cgaaaggggg	atgtgctgca	aggcgattaa	gttgggtaac	gccagggttt	tcccagtcac	4680
	gacgttgtta	aacgacggcc	agtgaattgt	aatacgactc	actatagggc	gaattggagc	4740
	tccaccgcgg	tgge					4754

65 <210> 81
 <211> 4773
 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pRK63pbsC31NLS-Re

5

<400> 81

10	ggcgcaccca	tgcccaagaa	gaagaggaag	gtgacacaag	gggttgtgac	cggggtggac	60
	acgtacggcg	gtgcttacga	ccgtcagtcg	cgcgagcgcg	agaattcgag	cgcagcaagc	120
	ccagcgacac	agcgtagcgc	caacgaagac	aaggcgcccg	accttcagcg	cgaagtcgag	180
	cgcgacgggg	gcccgttcag	gttcgtcggg	catttcagcg	aagcgccggg	cacgtcggcg	240
	ttcgggacgg	cggagcgccc	ggagttcgaa	cgcacccctga	acgaatgccg	cgcggggcgg	300
	ctcaacatga	tcattgtcta	tgacgtgtcg	cgcttcctgc	gcctgaaggc	catggacgcg	360
	attccgattg	tctcggaatt	gctcgccctg	ggcgtgacga	ttgtttccac	tcaggaaggc	420
15	gtcttcggcg	agggaaacgt	catggacctg	attcacctga	ttatgcggct	cgcgcgctcg	480
	cacaaagaat	cttcgctgaa	gtcggcgaag	attctcgaca	cgaagaacct	tcagcgcgaa	540
	ttgggggggt	acgtcggcgg	gaaggcgcc	tacggcttcg	agcttggttc	ggagacgaag	600
	gagatcacgc	gcaacggccg	aatgggtcaat	gtcgtcatca	acaagcttgc	gcactcgacc	660
	actcccctta	ccggaccctt	cgagttcgag	cccgacgtaa	tcgggtgggt	gtggcggtg	720
20	atcaagacgc	acaaacacct	tcctttcaag	ccgggcagtc	aagcgcccat	tcacccgggc	780
	agcatcacgg	ggctttgtaa	gcgcattggac	gctgacggcg	tgccgacccg	gggcgagacg	840
	attgggaaga	agaccgcttc	aagcgccctg	gaccggcga	ccgttatgcg	aatccttcgg	900
	gacccgcgta	ttcggggctt	cgcgcgtgag	gtgatctaca	agaagaagcc	ggacggcagc	960
	ccgaccacga	agattgaggg	ttaccgcatt	cagcgcgacc	cgatcacgct	ccggccggtc	1020
25	gagcttgatt	gcggaccgat	catcgagccc	gctgagtggt	atgagcttca	ggcgtgggtg	1080
	gacggcaggg	ggcgcggcaa	ggggctttcc	aagagcggaag	ccattctgtc	cgccatggac	1140
	aagctgtact	gcgagtgtgg	cgcgcgtcatg	acttogaagc	gcggggaaga	atcgatcaag	1200
	gactcttacc	gctgcgctcg	ccggaagggtg	gtcgaccogt	ccgcacctgg	gcagcacgaa	1260
	ggcacgtgca	acgtcagcat	ggcggcactc	gacaagttcg	ttgcggaacg	catcttcaac	1320
30	aagatcaggg	acgccgaagg	cgacgaagag	acgttgggcg	ttctgtggga	agccgcccga	1380
	cgcttcggca	agctcactga	ggcgccctgag	aagagcgggc	aacgggcgaa	ccttgttgcg	1440
	gagcgcggcg	acgcctgaa	cgcccttgaa	gagctgtacg	aagaccgcgc	ggcaggcgcg	1500
	tacgacggac	ccgttggcag	gaagcacttc	cggaagcaac	aggcagcgct	gacgctccgg	1560
	cagcaagggg	cggaagagcg	gcttgccgaa	cttgaagcgg	ccgaagcccc	gaagcttccc	1620
35	cttgaccaat	ggttccccga	agacgcggac	gctgacccga	ccggccctaa	gtcgtgggtg	1680
	gggcgcgcgt	cagtagacga	caagcgogtg	ttcgtcgggc	ttctcgtaga	caagatcggt	1740
	gtcacgaagt	cgactacggg	cagggggcag	tgaaacgccc	tcgagaagcg	cgcttcgata	1800
	acgtggggcg	agccgcggac	cgacgacgac	gaagacgacg	cccaggacgg	cacggaagac	1860
	gtagcggcgt	aggcggcgcc	cgggctcgag	ggggggcccc	gtacccagct	tttgttccct	1920
40	ttagttaggg	tttaatttcga	gcttggcgta	atcatggtoa	tagctgtttc	ctgtgtgaaa	1980
	ttgttatccg	ctcacaattc	cacacaacat	acgagccgga	agcataaagt	gtaaagcctg	2040
	gggtgcctaa	tgagttagct	aaotcacatt	cgctcactgc	cgcttttcca	ccgttttcca	2100
	gtcgggaaac	ctgtcgtgcc	agctgcatta	atgaatgggc	caacgcgcgg	ggagaggcgg	2160
45	tttgcgattt	gggcgctctt	ccgcttcttc	gctcaotgac	tcgctgcgct	cggtcggtcg	2220
	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	2280
	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	2340
	ggccgcgttg	ctggcggtttt	tcocataggt	ccgccccctc	gacgagcatc	acaaaaatcg	2400
	acgtcaagtg	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttccccc	2460
	tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgcgc	cttaccggat	acctgtccgc	2520
50	ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	2580
	ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccggtc	agcccgaccg	2640
	ctgcgcctta	tcgggtaact	atogtcttga	gtccaacccg	gtaagacacg	acttatcgcc	2700
	actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	2760
55	gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	2820
	tctgctgaag	ccagttacct	tcggaaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	2880
	caccgctggg	agcgggtgggt	tttttgggtt	caagcagcag	attacgcgca	gaaaaaaagg	2940
	atctcaagaa	gatcctttga	tcctttctac	ggggtctgac	gctcagtgga	acgaaaactc	3000
	acgttaaggg	atthttggta	tgagattatc	aaaaaggatc	ttcacctaga	tccttttaaa	3060
	ttaaaaatga	agtttttaaa	caatctaaag	tatatatgag	taaacttggg	ctgacagtta	3120
60	ccaatgctta	atcagtggag	caactatctc	agcatctgtg	ctatttcggt	catccatagt	3180
	tgctgacttc	cccgtcgtgt	agataactac	gatacgggag	ggcttaccat	ctggccccag	3240
	tgtgcaatg	ataccgcgag	acccacgctc	accggctcca	gatttatcag	caataaacca	3300
	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	ttatccgcct	ccatccagtc	3360
	tattaattgt	tgccgggaag	ctagagtaag	tagttcgcca	gttaaatagtt	tgcgcaacgt	3420
65	tgttgccatt	gctacaggca	tcgtgggtgc	acgctcgtcg	tttgggtatg	cttcctcag	3480
	ctcgggttcc	caacgatcaa	ggcgagttac	atgttcccc	atgttgtgca	aaaaagcggt	3540
	tagctccttc	ggtcctccga	tcgttgtcag	aagtaagttg	gcgcgagtg	tatcactcat	3600

75

	ggttatggca	gcactgcata	attctctttac	tgtcatgccca	tcogtaagat	gottttctgt	3660
	gactggtgag	tactcaacca	agtcattctg	agaatagtgt	atgogggcag	cgagttgctc	3720
	ttgcccggcg	tcaatacggg	ataataccgc	gccacatagc	agaactttta	aagtgtctcat	3780
5	cattggaaaa	cgttcttcgg	ggcgaaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	3840
	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	tcttttactt	tcaccagcgt	3900
	ttctgggtga	gcaaaaacag	gaaggcaaaa	tggcgcaaaa	aagggaataa	ggcgacacag	3960
	gaaatgttga	atactcatac	tcttcctttt	tcaatattat	tgaagcattt	atcaggggta	4020
	ttgtctcatg	agcggataca	tatttgaatg	tatttagaaa	aataaacaaa	taggggttcc	4080
10	gogcacattt	ccccgaaaag	tgccacctaa	attgtaagcg	ttaataattt	gttaaaattc	4140
	gogttaaaat	tttgtaaaat	cagctcattt	tttaaccaat	aggccgaaat	cggcaaaatc	4200
	ccttataaat	caaaagaata	gaccgagata	gggttgagtg	ttgttccagt	ttggaacaag	4260
	agtcactat	taaagaacgt	ggactccaac	gtcaaagggc	gaaaaaccgt	ctatcagggc	4320
	gatggccac	tacgtgaacc	atcaccttaa	tcaagttttt	tggggtcgag	gtgccgtaaa	4380
	gcactaaatc	ggaaccttaa	agggagcccc	cgatttagag	cttgacgggg	aaagccggcg	4440
15	aacgtggcga	gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	4500
	gtagcggta	cgctgcgcgt	aaccaccaca	cccgccgcgc	ttaatgcgcc	gctacagggc	4560
	gcgtccattt	cgccattcag	gctgcgcaac	tgttgggaaag	ggcgatcggt	gcgggcctct	4620
	tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggttaacg	4680
	ccagggtttt	cccagtcacg	acgttgtaaa	acgcggccca	gtgaattgta	atacgactca	4740
20	ctatagggcg	aattggagct	ccaccgcggt	ggc			4773

<210> 82

<211> 7803

25 <212> DNA

<213> Artificial Sequence

<220>

30 <223> Description of Artificial Sequence: vector
pPGKattA1

<400> 82

	tatcatgtct	ggatccgcgt	taacacctaa	gaaggcgaag	ttttccttac	accttgacaga	60
35	tataaagtgt	ctaacagttt	aaaatatccg	atgaggcata	tttatgttgg	acccgtagct	120
	cagccaggat	agagcactgg	cctccggagc	cggaggtccc	gggttcaaat	ccggcggggt	180
	ccgtatatta	ctttttgatt	cagattagat	ttgtaaatct	ttattacaag	gataatttga	240
	tcttgtatat	tggttaactct	ctactctata	atttttatga	gaaattcaca	gtcgtccctt	300
	tataccataa	atagctaagt	ttgtcaaagt	tcttattaaa	ctctccatgt	agagattaaa	360
40	tcggatccat	aacttcgtat	agcatacatt	atacgaagtt	ataccggggc	acctgggtcg	420
	cgagtgcctt	ggcactggcc	gtcgtttttac	aacgtcgtga	ctgggaaaac	cctggcggtta	480
	cccaacttaa	tcgccttgca	gcacatcccc	ctttcgccag	ctggcgtaat	agcgaagagg	540
	cccgccacga	tcgccccttc	caacagttgc	gcagcctgaa	tggcgaatgg	cgctttgcct	600
	ggtttccggc	accagaagcg	gtgccggaaa	gctggctgga	gtgcgatctt	cctgaggccg	660
45	atactgtcgt	cgtccctcca	aactggcaga	tgcacgggta	cgatgcgccc	atctacacca	720
	acgtaacctt	tcccattacg	gtcaatccgc	cgtttgttcc	cacgggagaat	ccgacggggt	780
	gttactgcgt	cacatttaac	gttgatgaaa	gctggctaca	ggaaggccag	acgcgaatta	840
	tttttgatgg	cgtttaactcg	gcgtttctac	ctttcgctcaa	cgggcgctgg	gtcggttacg	900
	gccaggacag	tcgtttgcgc	tctgaatttg	acctgagcgc	attttttacgc	gccggagaaa	960
50	accgcctcgc	ggtgatgggt	ctgcgttgga	gtgacggcag	ttatctggaa	gatcaggata	1020
	tgtggcggat	gagcggcatt	ttccgtgaag	tctcggtgct	gcataaaccg	actacacaaa	1080
	tcagcgattt	ccatggttgc	actcgcttta	atgatgattt	cagccgcgct	gtactggagg	1140
	ctgaagtcca	gatgtgcggc	gagttgcgtg	actacctacg	ggtaacagtt	tctttatggc	1200
	agggtgaaac	gcaggtcgcc	agcggcaccg	cgcttttcgg	cggtgaaatt	atcgatgagc	1260
55	gtggtgggta	tgccgatcgc	gtcacactac	gtctgaacgt	cgaaaacccg	aaactgtgga	1320
	gcgccgaaat	cccgaatctc	tatcgtgcgg	tggttgaact	gcacaccggc	gacggcacgc	1380
	tgattgaagc	agaagcctgc	gatgtcgggt	tccgcgaggt	gcggattgaa	aatgggtctgc	1440
	tgctgctgaa	cggcaagccg	ttgctgattc	gaggcggtta	ccgtcacgag	catcatcttc	1500
	tgcatggtca	ggtcatggat	gagcagacga	tggtgcagga	tatcctgctg	atgaagcaga	1560
60	acaactttta	cgccgtgcgc	tggttcgatt	atccgaacca	tccgctgtgg	tacacgctgt	1620
	gcgaccgcta	cggcctgtat	gtggtggatg	aagccaatat	tgaaacccac	ggcatgggtgc	1680
	caatgaatcg	tctgaccgat	gatccgcgct	ggctaccggc	gatgagcgaa	cgcgtaacgc	1740
	gaatggtgca	gcgcgatcgt	aatcacccga	gtgtgatcat	ctggctcgctg	gggaatgaat	1800
	cagggccagc	cgctaataac	gacgogctgt	atcgctggat	caaactctgtc	gacctctccc	1860
65	gcccggtgca	gctgtaaggc	ggcgagcgcg	acaccacggc	caccgatatt	attttcccga	1920
	tgtaacgcgc	cgtggatgaa	gaccagccct	tcccggttgt	gccgaaatgg	tccatcaaaa	1980
	aatggctttc	gctacctgga	gagacgcgcc	cgctgatcct	ttgcgaatac	gccacgcgca	2040
	tgggtaacag	tcttggcggt	tccgctaaat	actggcaggc	gtttcgtcag	tatccccggt	2100

	tacagggcgg	cttcgtctgg	gactgggtgg	atcagtcgct	gattaaatat	gatgaaaacg	2160
	gcaaccctgt	gtcggcttac	ggcgggtgatt	ttggcgatac	gccgaacgat	cgccagttct	2220
	gtatgaacgg	tctggctctt	gccgaccgca	cgcccgatcc	agcgtgacg	gaagcaaac	2280
5	accagcagca	gtttttccag	ttccgtttat	ccgggcaaac	catcgaagt	accagcgaat	2340
	acctgttccg	tcatagcgat	aacgagctcc	tgcactggat	ggtggcgctg	gatggtaagc	2400
	cgctggcaag	cggtgaagt	cctctggatg	tcgctccaca	aggtaaacag	ttgattgaac	2460
	tgcctgaact	accgcagccg	gagagcgccg	ggcaactctg	gctcacagta	cgcgtagtgc	2520
	aaccgaacgc	gaccgcatgg	tcagaagccg	ggcacatcag	cgcttggcag	cagtggcgctc	2580
10	tggcggaaaa	cctcagtgtg	acgctccccg	ccgcgtccca	cgccatcccg	catctgacca	2640
	ccagcgaaat	ggatttttgc	atcgagctgg	gtaataagcg	ttggcaattt	aaccgccagt	2700
	caggctttct	ttcacagatg	tggattggcg	ataaaaaaca	actgctgacg	ccgctgcgcg	2760
	atcagttcac	ccgtgcaccg	ctggataacg	acattggcgt	aagtgaagcg	accgcatttg	2820
	accctaacgc	ctgggtcgaa	cgctgggaag	cgccgggcca	ttaccaggcc	gaagcagcgt	2880
15	tgttgcaagt	cacggcagat	acacttgctg	atgcggtgct	gattacgacc	gctcacgcgt	2940
	ggcagcatca	ggggaaaacc	ttatttatca	gccggaaaac	ctaccggatt	gatggtagt	3000
	gtcaaatggc	gattaacgtt	gatgttgaag	tggcgagcga	tacaccgcat	ccggcgcgga	3060
	ttggcctgaa	ctgccagctg	gcgcaggtag	cagagcgggc	aaactggctc	ggattagggc	3120
	cgcaagaaaa	ctatcccgac	cgcttactgt	cgccgtgttt	tgaccgctgg	gatctggcgt	3180
20	tgtcagacat	gtataccccc	tacgtcttcc	cgacgcaaaa	cggtctgcgc	tgcgggacgc	3240
	gcgaattgaa	ttatggccca	caccagtggc	gcggcgactt	ccagttcaac	atcagccgct	3300
	acagtcaaca	gcaactgatg	gaaaccagcc	atcgccatct	gctgcacgcg	gaagaaggca	3360
	catggctgaa	tatcgacggg	ttccatatgg	ggattggtgg	cgacgactcc	tggagcccg	3420
	cagtatccgg	ggaattccag	ctgagcgccg	gtcgctacca	ttaccagttg	gtctggtgtc	3480
25	aaaaataata	ataaccgggc	aggggggac	tttgtgaagg	aaccttactt	ctgtggtgtg	3540
	acataattgg	acaaactacc	tacagagatt	taaagctcta	aggtaaatat	aaaattttta	3600
	agtgtataat	gtgttaaaat	actgattcta	attgtttgtg	tatttttagat	tccaacctat	3660
	ggaactgatg	aatgggagca	gtggtggaat	gccagatcca	gacatgataa	gatacattga	3720
	tgagtttgga	caaaccacaa	ctagaatgca	gtgaaaaaaa	tgctttat	gtgaaatttg	3780
	tgatgctatt	gcctttat	taaccattat	aagctgcaat	aaacaagtta	acaacaacaa	3840
30	ttgcattcat	tttatgtttc	aggttcaggg	ggaggtgtgg	gaggtttttt	aaagcaagta	3900
	aaacctctac	aaatgtggta	tggctgatta	tgatctgcgg	ccgcagggcc	tcgtgatacg	3960
	cctattttta	taggttaatg	tcatgataat	aatggtttct	tagacgtcag	gtggcacttt	4020
	tcggggaaat	gtgcgcggaa	ccctattttg	tttatttttc	taaatacatt	caaataatga	4080
	tcgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	4140
35	gagtattcaa	catttccgtg	tcgcccttat	tccttttttt	gcggcatttt	gccttccctg	4200
	ttttgctcac	ccagaaaacg	tggtgaaagt	aaaagatgct	gaagatcagt	tgggtgcacg	4260
	agtgggttac	atcgaactgg	atctcaacag	cggttaagatc	cttgagagtt	ttcgccccga	4320
	agaacgtttt	ccaatgatga	gcacttttaa	agttctgcta	tgtggcgcg	tattatcccg	4380
40	tattgacgcc	gggcaagagc	aactcgggtc	ccgcatacac	tattctcaga	atgacttggt	4440
	tgagtactca	ccagtcacag	aaaagcatct	tacggatggc	atgacagtaa	gagaattatg	4500
	cagtgcctgc	ataaccatga	gtgataacac	tcgggccaac	ttacttctga	caacacagcg	4560
	aggaccgaag	gagctaaccg	cttttttgca	caacatgggg	gatcatgtaa	ctcgccctga	4620
	tcgttgggaa	ccggagctga	atgaagccat	accaaaccgac	gagcgtgaca	ccacgatgcc	4680
45	tgtagcaatg	gcaacaacgt	tgcgcaaac	attaactggc	gaactactta	ctctagcttc	4740
	ccggcaacaa	ttaatagact	ggatggaggc	ggataaagtt	gcaggaccac	ttctgcgctc	4800
	ggcccttccg	gctggctggt	ttattgctga	taaacttgga	gccggtgagc	gtgggtctcg	4860
	cggtatcatt	gcagcactgg	ggccagatgg	taaagcctcc	cgtatcgtag	ttatctacac	4920
	gacggggagt	caggcaacta	tggatgaacg	aaatagacag	atcgctgaga	taggtgcctc	4980
50	actgattaag	cattggtaac	tgtcagacca	agtttactca	tatatacttt	agattgattt	5040
	aaaacttcac	ttttaattta	aaaggatcta	ggtgaagatc	ctttttgata	atctcatgac	5100
	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gaocccgtag	aaaagatcaa	5160
	aggatcttct	tgagatcctt	tttttctgcg	cgtaactctg	tgcttgcaaa	caaaaaaac	5220
	accgctacca	cggtggtttt	gtttgccgga	tcaagagcta	ccaactcttt	ttccgaaggt	5280
55	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	cgtagttagg	5340
	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	gctctgctaa	tcctgttacc	5400
	agtggctgct	gccagtggcg	ataagtcgtg	tcttacccgg	ttggactcaa	gacgatagtt	5460
	accggataag	gcgcagcggt	cggtgtgaac	ggggggttcg	tgcacacagc	ccagcttggg	5520
	tccgaagagg	tacacogaac	tgagatacct	acagcgtgag	ctatgagaaa	gcgccacgct	5580
60	cacgagggag	cttccagggg	gaaacgcctg	ggtaagcggc	agggtcggaa	caggagagcg	5640
	cctctgactt	gagcgtcgat	ttttgtgatg	gtatctttat	agtccgtgct	ggtttcgcca	5700
	cgccagcaac	gcggcctttt	tacggttcc	ctcgtcaggg	gggaggagcc	tatggaaaaa	5760
	ctttcctgcg	ttatccctcg	attctgtgga	ggccttttgc	tggccttttg	ctcacatgtt	5820
	taccctctgc	cgcacccgaa	cgaccagcgc	taaccgtatt	accgcctttg	agtgaagctga	5880
65	gcgcccaata	cgcaaacggc	ctctcccgcg	cagcagtgca	gtgagcaggg	aagcgggaaga	5940
	cgacaggttt	cccgactgga	aagcgggcag	gcgttggccg	attcattaat	gcagctggca	6000
	cactcattag	gcaccccagg	ctttacactt	tgagcgcaac	gcaattaatg	tgagttagct	6060
				tatgcttccg	gctcgtatgt	tgtgtggaat	6120

5 tgtgagcggg taacaatttc acacaggaaa cagctatgac catgattacg ccaagctggc 6180
 gcgtcgaggga attctaccgg gtaggggagg cgcttttccc aaggcagctc ggagcatgcg 6240
 ctttagcagc cccgctgggc acttggcgct acacaagtgg cctctggctc gcacacattc 6300
 cacatccacc ggtaggcgcc aaccggctcc gttctttggg ggccctctcg cgcacacctc 6360
 tactctctcc ctagtacagg agttcccccc cgcgccgcag ctgcgctcgt gcaggacgtg 6420
 acaaatggaa gtagcacgtc tcaactagtct cgtgcagatg gacagcaccg ctgagcaatg 6480
 gaagcgggta ggcctttggg gcagcggcca atagcagctt tgctccttcg ctttctgggc 6540
 tcagaggctg ggaaggggtg ggtccggggg cgggctcagg ggcgggctca gggcgggggc 6600
 10 gggcgcccca aggtcctccg gaggcccgcc attctgcacg cttcaaaagc gcacgtctgc 6660
 cgcgctgttc tctctcttcc catctccggg cctttcgacc tgcagcccgg tacagttcga 6720
 ataacttcgt atagcataca ttatacgaag ttataagctt gcatgcctgc aggtcggccg 6780
 ccacgaccgg ccggccgggtg ccgccaccat cccctgacct acgcccctga cccctcacia 6840
 ggagacgacc ttccatgacc gagtacaagg ccacggtgcg cctcgccacc cgcgacgacg 6900
 tccccgggca cgtacgcacc ctgcgcgcgc cgttcgcca ctaccccgcc acgcgccaca 6960
 15 ccgtcgaccc ggaccgccac atcgagcggg tcaccgagct gcaagaactc ttctctcagc 7020
 gcgtcgggct cgacatcggc aaggtgtggg tcggcgacga cggcgccgcg gtggcggtct 7080
 ggaccacgcc ggagagcgtc gaagcggggg cgggtgttcg cgagatcggc ccgcgcatgg 7140
 ccgagttgag cggttcccgg ctggccgcgc agcaacagat ggaaggcctc ctggcgccgc 7200
 accggcccaa ggagcccgcg tggttcctgg ccacgctcgg cgtctcgccc gaccaccagg 7260
 20 gcaagggtct gggcagcgcc gtctgtctcc ccgagtgga ggcgcccgag cgcgcggggg 7320
 tgcccgctt cctggagacc tccgcgcccc gcaacctccc cttctacgag cggctcggct 7380
 tcaccgtcac cgcgacgtc gagtgcgccg aggaccgcgc gacctggtgc atgaccgcga 7440
 agcccggtgc ctgacgcccg cccacgacc cgcagcgccc gaccgaaagg agcgcacgac 7500
 cccatggctc cgaccgaagc cgaccgggc ggcccgccg accccgcacc cgcgcccgag 7560
 25 gcccacgac tctagaggat cataactcag cataccacat ttgtagagg tttacttgc 7620
 ttaaaaaacc tcccacacct cccctgaac ctgaaacata aaatgaatgc aattgttgtt 7680
 gttaacttgt ttattgcagc ttataatggt tacaataaaa gcaatagcat cacaatttc 7740
 acaataaag cattttttc actgcattct agttgtggtt tgtccaaact catcaatgta 7800
 30 tct 7803

<210> 83

<211> 8167

<212> DNA

35 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pPGKattA2

40

<400> 83

tatcatgtct ggatccgctg taacacctaa gaaggcgaag ttttctttac accttgcaga 60
 tataaagtgt ctaacagttt aaaatatccg atgaggcata tttatgttgg acccgtagct 120
 45 cagccaggat agagcactgg cctccggagc cggaggtccc gggttcaaat cccggcgggt 180
 cgttatatta ctttttgatt cagattagat ttgtaaatct ttattacaag gataatttga 240
 tcttgtatat tggtaactct ctactctata atttttatga gaaattcaca gtcgtccctt 300
 tataccataa atagctaagt ttgtcaaagt tcttattaaa ctctccatgt agagattaaa 360
 toggatccat aacttcgtat agcatacatt atacgaagtt ataccgggcc accatggtcg 420
 50 cgagtagctt ggcaactggc gtctgtttac aacgtcgtga ctgggaaaac cctggcgtta 480
 cccaacttaa tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg 540
 ccgcgaccga tcgccttcc caacagttgc gcagcctgaa tggcgaatgg cgttttgcct 600
 ggtttccggc accagaagcg gtgccggaaa gctggctgga gtgcgatctt cctgaggccg 660
 atactgtcgt cgtccctca aactggcaga tgcacggtta cgtgcgccc atctacacca 720
 55 acgtaacctc toccattacg gtcaatccgc cgtttgttcc caccgagaat ccgacgggtt 780
 gttactcgtc cacatttaat gttgatgaaa gctggctaca ggaaggccag acgcgaatta 840
 tttttgatgg cgttaactcg gcgtttcatc tgtggtgcaa cggcgctgg gtccggttacg 900
 gccaggacag tcgtttgccc tctgaatttg acctgagcgc atttttacgc gccggagaaa 960
 accgcctcgc ggtgatggtg ctgcgttgga gtgacggcag ttatctggaa gatcaggata 1020
 tgtggcggat gagcggcatt ttccgtgacg tctcgttgct gcataaaccg actacacaaa 1080
 60 tcagcgatctt ccattgtgcc actcgttata atgatgatt cagccgcgct gtactggagg 1140
 ctgaagttca gatgtgcggc gaggttgcgt actacctacg ggtaacagtt tctttatggc 1200
 agggtgaaac gcaggtcgcc agcggcaccg cgcctttcgg cggtgaaatt atcgatgagc 1260
 gtggtgggta tgccgatcgc gtcacactac gtctgaacgt cgaaaaaccg aaactgtgga 1320
 gcgcggaaat cccgaatctc tatcgtgcgg tggttgaact gcacaccgcc gacggcacgc 1380
 65 tgattgaagc agaagcctgc gatgtcggtt tccgcgaggt gcggattgaa aatggtctgc 1440
 tgotgtgaa cggcaagccg ttgctgattc gagcggttaa ccgtcacgag catcatctc 1500
 tgcattggtc ggtcatggtg gaggcagcga tgggtgcagg tatctcgtcg atgaagcaga 1560

	acaacttttaa	cgccgtgccc	tggttcgcatt	atccgaacca	tccgtgtgtg	tacacgctgt	1620
	gogaccgcta	cgccctgtat	gtggtggatg	aagccaatat	tgaaacccac	ggcatggtgc	1680
	caatgaatcg	tctgaccgat	gatccgcgct	ggctaccggc	gatgagcgaa	cgcgtaacgc	1740
5	gaatgggtgca	gcgcgatcgt	aatcaccocga	gtgtgatcat	ctggtoctgt	gggaatgaat	1800
	caggccacgg	cgctaatac	gacgcgctgt	atcgctggat	caaatactgt	gatccttccc	1860
	gcccgggtgca	gtatgaagcg	ggcggagccg	acaccacggc	cacogatatt	atttgcccg	1920
	tgtacgcgcg	cgtggatgaa	gaccagccct	tcccgctgt	gocgaaatgg	tccatcaaaa	1980
	aatggctttc	gctacctgga	gagacgcgcc	cgctgatcct	ttgogaatac	gcccacgcga	2040
10	tgggtaacag	tcttggcggt	ttcgctaaat	actggcagcg	gtttcgtcag	tatccccgtt	2100
	tacagggcg	cttcgtctgg	gactgggtgg	atcagtcgct	gattaaatat	gatgaaaacg	2160
	gcaacccgtg	gtcggcttac	ggcgggtgatt	ttggcgatac	gocgaacgat	cgccagtctc	2220
	gtatgaacgg	tctggtcttt	gocgacgcga	cgcgcatacc	agocgtgacg	gaagcaaaaa	2280
	accagcagca	gtttttccag	ttccgtttat	cggggcaaac	catogaagtg	accagcgaat	2340
15	acctgttccg	tcatagcgat	aacgagctcc	tgcactggat	ggtggcgctg	gatggtaagc	2400
	cgctggcaag	cgggtgaagt	cctctggatg	tgcctccaca	aggtaaacag	ttgattgaac	2460
	tgcctgaact	accgcagccg	gagagcgccg	ggcaactctg	gotcacagta	cgcgtagtgc	2520
	aacogaacgc	gaccgcagtg	tcagaagccg	ggcacatcag	cgccctggcg	cagtggcgte	2580
	tggcggaaaa	cctcagtgtg	acgctccccg	cgcgctccca	cgccatcccg	catctgacca	2640
20	ccagcgaaat	ggatttttgc	atcgagctgg	gtaataagcg	ttggcaattt	aaccgccagt	2700
	caggctttct	ttcacagatg	tggattggcg	ataaaaaaca	actgctgacg	ccgctgcgcg	2760
	atcagttcac	cogtgcaccg	ctggataacg	acattggcgt	aagtgaagcg	acccgcattg	2820
	accctaacgc	ctgggtcgaa	cgctggaagg	cggcgggcca	ttaccaggcc	gaagcagcgt	2880
	tgttgacagt	cacggcagat	acacttgctg	atcggtgtct	gattacgacc	gctcacgcgt	2940
25	ggcagcatca	ggggaaaacc	ttattttatca	gcccgaaaaac	ctacoggatt	gatggtagt	3000
	gtcaaatggc	gattaccgtt	gatgttgaag	tggcgagcga	tacaccgcat	ccggcgcgga	3060
	ttggcctgaa	ctgccagctg	gcccaggtag	cagagcgggt	aaactggctc	ggattagggc	3120
	cgcaagaaaa	ctatcccgac	cgccctactg	ccgcctgttt	tgacogctgg	gatctgccat	3180
	tgtcagacat	gtataccccg	tacgtcttcc	cgagcgaaaa	cggtctgcgc	tgcgggacgc	3240
30	gogaattgaa	ttatggccca	caccagtggc	gcccgcactt	ccagttcaac	atcacccgct	3300
	acagtcaaca	gcaactgatg	gaaaccagcc	atcgccatct	gctgcacgcg	gaagacggca	3360
	catggctgaa	tatcgacggg	ttccatatgg	ggatttgggtg	cgacgactcc	tggagcccg	3420
	cagtatcgcc	ggaattccag	ctgagcgccg	gtcgctacca	ttaccagttg	gtctggtgtc	3480
	aaaaataata	ataaccgggc	aggggggatc	tttgtgaagg	aaccttactt	ctgtggtgtg	3540
35	acataattgg	acaaactacc	tacagagatt	taaagctcta	aggtaaatat	aaaattttta	3600
	agtgtataat	gtgtttaaact	actgattcta	attgtttgtg	tatttttagat	tccaacctat	3660
	ggaactgatg	aatgggagca	gtggtggaat	gccagatcca	gacatgataa	gatacattga	3720
	tgagtttggg	caaaccacaa	ctagaatgca	gtgaaaaaaa	tgctttattt	gtgaaatttg	3780
40	tgatgctatt	gctttatttg	taaccattat	aagctgcaat	aaacaagtta	acaacaacaa	3840
	ttgcattcat	tttatgtttc	aggttcaggg	ggaggtgtgg	gaggtttttt	aaagcaagta	3900
	aaacctctac	aaatgtggtg	tggctgatta	tgatctggcg	ccgcagggcc	tcgtgatacg	3960
	cctattttta	taggttaatg	tcattgataat	aatggtttct	tagacgtcag	gtggcacttt	4020
	tccgggaaat	gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaataatgta	4080
45	tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	4140
	gagtattcaa	catttccgtg	tccgcttat	tccctttttt	gcccgaattt	gccttccgtg	4200
	ttttgtctac	ccagaaacgc	tgggtgaaagt	aaaagatgct	gaagatcagt	tgggtgcacg	4260
	agtgggttac	atcgaactgg	atctcaacag	cggtaagatc	cttgagagtt	ttcgcccgca	4320
	agaacgtttt	ccaatgatga	gcacttttaa	agttctgcta	tgtggcgcg	tattatcccg	4380
	tattgacgcc	gggcaagagc	aactcggtcg	ccgcatacac	tattctcaga	atgacttggg	4440
50	tgagtactca	ccagtcacag	aaaagcatct	tacggatggc	atgacagtaa	gagaattatg	4500
	cagtgtgcc	ataaccatga	gtgataaac	tgcggccaac	ttacttctga	caacgatcgg	4560
	aggaccgaag	gagctaaccg	ctttttttgca	caacatgggg	gatcatgtaa	ctcgccctga	4620
	togttgggaa	cggagctga	atgaagccat	accaaacgac	gagcgtgaca	ccacgatgcc	4680
	tgtagcaatg	gcaacaacgt	tgcgcaaaact	attaactggc	gaactactta	ctctagcttc	4740
55	ccggcaacaa	ttaatagact	ggatggaggg	ggataaagtt	gcaggaccac	ttctgcgctc	4800
	ggcccttccg	gctggctggg	ttattgctga	taaactctgga	gcccgtgagc	gtgggtctcg	4860
	cggtatcatt	gcagcactgg	ggccagatgg	taagccctcc	cgtatcgtag	ttatctacac	4920
	gacggggagt	caggcaacta	tggatgaacg	aaatagacag	atcgctgaga	taggtgcctc	4980
	actgattaa	catttgtaac	tgtcagacca	agtttactca	tatatacttt	agattgattt	5040
60	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	ctttttgata	atctcatgac	5100
	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gaccccgtag	aaaagatcaa	5160
	aggatcttct	tgagatcctt	ttttctcgcg	cgtaactctg	tgcttgcaaa	caaaaaaacc	5220
	accgctacca	gcggtggttt	gtttgcggga	tcaagagcta	ccaactcttt	ttccgaaggt	5280
	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	cgtagttagg	5340
	ccaccacttc	aagaactctg	tagcacccgc	tacatacctc	gctctgctaa	tcctgttacc	5400
65	agtggctgct	gccagtggcg	ataagtcgtg	tcttacccgg	ttggactcaa	gacgatagtt	5460
	accggataag	gcccagcggt	cgggctgaac	ggggggttcg	tgcacacagc	ccagcttgga	5520
	gogaacgacc	tacaccgaac	tgagatacct	acagcgtgag	ctatgagaaa	gcgccacgct	5580

79

tcccgaggga agaaaggcgg acagggtatcc ggtaagcggc aggggtcggaa caggagagcg 5640
 caccgaggga cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca 5700
 cctctgactt gagcgtcgat ttttgtgatg ctctgcaggg gggcggagcc tatggaaaaa 5760
 5 cgcagcaaac ggggcctttt tacgggttcc ttacgttttg tggccttttg ctcacatgtt 5820
 ctttcctgcg ttatccctcg attctgtgga taaccgtatt accgcctttg agtgagctga 5880
 taccgctcgc cgcagccgaa cgaccgagcg cagcagtgca gtgagcagag aagcgggaaga 5940
 gcgccaata cgcacaaccg ctctcccccg gcgttgcccg attcattaat gcagctggca 6000
 cgacaggttt cctagtcagg aagcgggcag tgagcgcac gcaattaatg tgagttagct 6060
 cactcattag gcaccccgag ctttacactt tatgcttccg gctcgtatgt tgtgtggaat 6120
 10 tgtgagcggg taacaatttc acacaggaaa cagctatgac catgattacg ccaagctggc 6180
 gcgtcgagga attctaccgg gtaggggagg cgcttttccc aaggcagctt ggagcatgcg 6240
 ctttagcagc cccgctgggc acttgccgct acacaagtgg cctctggctc gcacacattc 6300
 cacatccacc ggtaggcgcc aaccggctcc gttcttttgg gtccccttcg cgccaccttc 6360
 tactcctccc ctatgcagg agttccccc cgcctcgag ctccgctcgt gcaggacgtg 6420
 15 acaaatggaa gtagcagtc tcaactagtct cgtgcagatg gacagcaccg ctgagcaatg 6480
 gaagcgggta ggccttttgg gcagcggcca atagcagctt tgctccttcg ctttctgggc 6540
 tcagaggctg ggaaggggtg ggtccggggg cgggctcagg ggccgggtca ggggcccggc 6600
 gggcggccga aggtcctccg gaggcccgcc attctgcacg cttcaaaagc gcacgtctgc 6660
 20 cgcgtctgtt tctcttctc catctccggg ccttttcgac tgacgcccgg tacagttcga 6720
 aggatccgag ttaacacctt agaagggcaa gttttcctta caccttgacg atataaagt 6780
 tctaacagtt taaaatatcc gatgaggcat atttatgttg gacccttagc tcagccagga 6840
 tagagcactg gcctccggag ccggagggtcc cgggttcaaa tcccggcggg tccgtatatt 6900
 actttttgat tcagattaga tttgtaaaac tttattacaa ggataatttg atcttgtata 6960
 25 ttggttaact totactctat aattttttat agaaattcac agtcgtccct ttataccata 7020
 aatagctaag tttgtcaaag ttcttattaa actctccatg tagagattaa atcgatcct 7080
 tcgaataact togtatagca tacattatac gaagttataa gcttgcatgc ctgcaggtcg 7140
 gccgccacga cggccgggcc ggtgccgcca ccatccctg acccagccc ctgacccctc 7200
 acaaggagac gaccttccat gaccgagtag aagcccacgg tgccgctcgc caccgcgcac 7260
 30 gacgtccccc gggccgtagc caccctcgcc gccggttcg ccgactaccc cgcacgcgc 7320
 cacaccgtcg acccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttctc 7380
 accgcgctcg ggctcgacat cggcaagggtg tgggtcgccg acgacggcgc cgcggtggcg 7440
 gtctggacca cgcgggagag cgtcgaagcg gggcggtgt tcgcccagat cggcccgcg 7500
 atggccgagt tgagcgggtc ccggttgccc gcgcagcaac agatggaagg cctcctggcg 7560
 35 ccgacccggc ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac 7620
 cagggcaagg gtctgggag cgcctcgtg ctccccggag tggaggcggc cgagcgcgc 7680
 ggggtgccc ccttcttgga gacctccg ccccgcaacc tccccttcta cagcggctc 7740
 ggcttcacgc tcaccggcga cgtcagtgcc ccgaaggacc gcgcgacctg gtgcattgac 7800
 cgcaagcccg gtgcctgacg cccgccccac gaccgcgacg gcccgaccga aaggagcgca 7860
 40 cgaccccatg gctccgaccg aagccgaccc gggcggcccc gccgaccccg cacccgcgcc 7920
 cgaggccac cgactctaga ggatcataat cagccatacc acatttttag aggttttact 7980
 tgctttaaaa aacctcccac acctccccct gaacctgaaa cataaaatga atgcaattgt 8040
 tgttgttaac ttgtttattg cagcttataa tggttacaaa taaagcaata gcatacaaa 8100
 tttcacaaat aaagcatttt ttctactgca ttctagttgt ggtttgtoca aactcatcaa 8160
 45 tgtatct 8167

<210> 84

<211> 51

<212> DNA

50 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer XisA1

55 <400> 84

ataagaatgc ggccgcccga tatgcaaaat cagggtcaag acaaatatca a

51

<210> 85

60 <211> 76

<212> DNA

<213> Artificial Sequence

<220>

65 <223> Description of Artificial Sequence: primer XisA2

<400> 85

80

ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgcaaaatc aggggtcaaga 60
 caaatatcaa caagcc 76

5 <210> 86
 <211> 44
 <212> DNA
 <213> Artificial Sequence

10 <220>
 <223> Description of Artificial Sequence: primer XisA3

<400> 86
 ataagaatgc ggccgctcaa ctattcttat aagctatttc catc 44

15 <210> 87
 <211> 82
 <212> DNA
 20 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer nifD1

25 <400> 87
 cgatggctct tcccttccgt caaatgcact cttgggatta ctccgaacct agcgatgggg 60
 tgcaaagtgc agatcagata ag 82

30 <210> 88
 <211> 82
 <212> DNA
 <213> Artificial Sequence

35 <220>
 <223> Description of Artificial Sequence: primer nifD2

<400> 88
 cgcttatctg atctgacatt tgcaccccat cgctagggtc ggagtaatcc caagagtgc 60
 40 ttgacggaa gggaagagcc at 82

<210> 89
 <211> 74
 45 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer nifD3

50 <400> 89
 gatcagctgt tgaaagctat taaaccacaa aaaggattac tccggccctt atcacggtta 60
 cgacggattt gcta 74

55 <210> 90
 <211> 74
 <212> DNA
 <213> Artificial Sequence

60 <220>
 <223> Description of Artificial Sequence: primer nifD4

<400> 90
 65 gatctagcaa atccgtcgta accgtgataa gggccggagt aatccttttt gtggttta 60
 agctttcaac agct 74

<210> 91
 <211> 41
 <212> DNA
 5 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: primerSSV1-1

 10 <400> 91
 ataagaatgc ggccgcccga tatgacgaaa gataagaagc g 41

 <210> 92
 15 <211> 70
 <212> DNA
 <213> Artificial Sequence

 <220>
 20 <223> Description of Artificial Sequence: primer SSV1-2

 <400> 92
 ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgacgaaag ataagacgcg 60
 ttataaatac 70
 25

 <210> 93
 <211> 47
 <212> DNA
 30 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: primer SSV2

 35 <400> 93
 tgtccccgggc tcgaaaccgg ggggatccgc ttgtagggga gtatccc 47

 <210> 94
 40 <211> 47
 <212> DNA
 <213> Artificial Sequence

 <220>
 45 <223> Description of Artificial Sequence: primer SSV3

 <400> 94
 gagccccggga caagcgggaag cgggtggtgga aaagagggaa ctgaacg 47
 50

 <210> 95
 <211> 33
 <212> DNA
 <213> Artificial Sequence
 55

 <220>
 <223> Description of Artificial Sequence: primer SSV4

 <400> 95
 60 atcgctcgag tcagaccctt tttagccatt ccg 33

 <210> 96
 <211> 40
 65 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer SSV5

<400> 96

5 atcgttcgaa ggatccgcgt taacacctaa gaaggcgaag

40

<210> 97

<211> 38

10 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer SSV6

15

<400> 97

atcgttcgaa ggatccgatt taatctctac atggagag

38

20 <210> 98

<211> 64

<212> DNA

<213> Artificial Sequence

25 <220>

<223> Description of Artificial Sequence: primer C31-2

<400> 98

30 ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgacacaag gggttgtgac 60
cggg 64

<210> 99

<211> 4831

35 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector pRK41

40

<400> 99

ggccgcacca tgcccaagaa gaagaggaag gtgacacaag gggttgtgac cggggtggac 60
 acgtacgcgg gtgcttacga ccgtcagtcg ccgagcgcg agaattcgag cgcagcaagc 120
 ccagcgacac agcgtagcgc caacgaagac aaggcggccg accttcagcg cgaagtcgag 180
 45 cgcgacgggg gccggttcag gttcgtcggg catttcagcg aagcgccggg cagtcggcg 240
 ttccgggacgg cggagcgccc ggagttcgaa cgcatactga acgaatgccg cgcggggcg 300
 tcaaacatga tcattgtcta tgacgtgtcg gccttctcgc gcctgaaggc catggacgcg 360
 attccgattg tctcggaatt gctcgccctg ggcgtagcga ttgtttccac tcagggaaggc 420
 gtcttccggc agggaaacgt catggacctg attcaactga ttatgcggct cgacgcgtcg 480
 50 cacaagaat cttcgctgaa gtggcggaag attctcgaca cgaagaacct tcagcgcgaa 540
 ttgggcgggg acgtcggcgg gaaggcgcc taccggttcg agcttggttc ggagacgaag 600
 gagatcacgc gcaacggccg aatgggtcaat gtcgtcatca acaagcttgc gcactcgacc 660
 actcccctta ccggaccctt cgagttcgag ccgacgtaa tccggtggtg gtggcgtag 720
 atcaagacgc acaaacacct tcccttcaag ccgggcagtc aagccgccat tcaccggggc 780
 55 agcatcacgg ggctttgtaa ggcgatggac gctgacgcg tgccgacccg gggcgagacg 840
 attgggaaga agaccgcttc aagcgccctg gacccgcaa ccgttatgag aatccttcgg 900
 gacccgcgta ttgcgggctt cgcgctgag gtgatctaca agaagaagcc ggacggcacg 960
 ccgaccacga agattgaggg ttaccgcatt cagcgcgacc cgatcacgct ccggccggtc 1020
 gagcttgatt gcggaccgat catcgagccc gctgagtggt atgagcttca ggcgtgggtg 1080
 60 gacggcaggg ggcgcgcaa ggggctttcc cgggggcaag ccattctgtc cgccatggac 1140
 aagctgtact gcgagtgtgg cgcgctcatg acttogaagc gcggggaaga atcgatcaag 1200
 gactcttacc gctgccgtcg ccggaagggt gtcgacccgt ccgcacctgg gcagcacgaa 1260
 ggcacgtgca acgtcagcat ggcggaactc gacaagttcg ttgcggaaacg catcttcaac 1320
 aagatcaggg acgccgaagg cgaagaagag acgttggcgc ttctgtggga agccgccga 1380
 65 cgcttcggca agctcactga ggcgccctgag aagagcgggc aacgggcgaa ccttgttgcg 1440
 gagcgcgccg acgcctgaa cgccttgaag gagctgtacg aagaccgcgc ggcaggcgcg 1500
 tacgacggac ccgttggcag gaagcacttc ccggaagcaac aggcagcgct gacgctccgg 1560

	cagcaaggagg	cggaagagcg	gcttgccgaa	cttgaagccg	ccgaagcccc	gaagcttccc	1620
	cttgaccaat	ggttccccga	agacgccgac	gctgaccoga	ccggccctaa	gtcgtggtgg	1680
	gggcgcgogt	cagtagacga	caagcgcggtg	ttcgtcgggc	tcttcgtaga	caagatcggt	1740
5	gtcacgaagt	cgactacggg	cagggggcag	ggaacgcca	tcgagaagcg	cgcttcgac	1800
	acgtgggoga	agccggccgac	cgacgacgac	gaagacgacg	cccaggacgg	cacggaagac	1860
	gtagcggcgt	agcggccgct	ctagaactag	tggatcccc	gggctgcagg	aattcgatat	1920
	caagcttate	gataccgtcg	acctcgaggg	ggggcccggt	accagcttt	tgttcccttt	1980
	agtgaggggt	aatttcgagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	2040
10	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	aaagcctggg	2100
	gtgcctaata	agtgaagtaa	ctcacattaa	ttgcgtttcg	ctcactgccc	gctttccagt	2160
	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgcca	acgcgcgggg	agaggcgggt	2220
	tgcgtattgg	gcgtctcttc	gcttcctcgc	tcactgactc	gctgcgctcg	gtcgttcggc	2280
	tgcggcgagc	ggatcaagct	cactcaaagg	cggtaatatc	gttatccaca	gaatcagggg	2340
15	ataacgcagc	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccaggaa	cgtaaaaaag	2400
	ccgogttgct	ggcgtttttc	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	2460
	gctcaagtca	gaggtggcga	aaccggacag	gactataaag	ataccaggcg	tttccccctg	2520
	gaagctccct	cgtgcgctct	cctgttccga	ccctgcgct	taccggatac	ctgtccgct	2580
	ttctcccttc	gggaagcgtg	gcgtttcttc	atagctcacg	ctgtagggtat	ctcagttcgg	2640
20	tgtaggctgt	tcgtcccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaccgct	2700
	gcgccttate	cggttaactat	cgtcttgagt	ccaaccgggt	aagacacgac	ttatcgccac	2760
	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggg	gctacagagt	2820
	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggg	atctgcgctc	2880
	tgotgaagcc	agttacottc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	2940
25	ccgctggtag	cggtgggttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	3000
	ctcaagaaga	tcctttgagc	ttttctacgg	ggtctgacgc	tcagtggaa	gaaaactcac	3060
	gttaagggat	tttggtcag	agattatcaa	aaaggatctt	cacctagatc	cttttaaaatt	3120
	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtc	gacagttacc	3180
	aatgcttaat	cagtgaaggca	cctatctcag	cgatctgtct	atttcgttca	tccatagttg	3240
30	cctgaactccc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	ggccccagtg	3300
	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	ataaaccage	3360
	cagccgggaag	ggccgagcgc	agaagtgtgc	ctgcaacttt	atccgcctcc	atccagctca	3420
	tttaattggtg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	cgcaacggtg	3480
	ttgccattgc	tacaggcatc	gtgggtgtc	gctcgtcggt	tggtatggct	tcattcagct	3540
35	ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	aaagcgggta	3600
	gctccttcgg	tcctccgac	gttgtcagaa	gtaagttggc	cgcagtgta	tcactcatgg	3660
	ttatggcagc	actgcataat	tctcttact	tcattgccatc	cgtaagatgc	ttttctgtga	3720
	ctggtgagta	ctcaaccag	tcattctag	aatagtgtat	gcggcgacog	agttgctctt	3780
	gccccggcgc	aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	3840
40	ttggaaaacg	ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	agatccagtt	3900
	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	accagcgttt	3960
	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	gcgacacgga	4020
	aatgttgaat	actcatactc	ttcctttttc	aatattattg	aagcattttat	cagggttatt	4080
45	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	ggggttccgc	4140
	gcacatttcc	ccgaaaagtg	ccacctaaat	tgtaagcggt	aatattttgt	taaaattcgc	4200
	gttaaatattt	tgtaaatca	gctcattttt	taaccaatag	gccgaaatcg	gcaaaatccc	4260
	ttataaatca	aaagaataga	ccgagatagg	gttgagtgtt	gttccagttt	ggaacaagag	4320
	tcactatta	aagaacgtgg	actccaacgt	caaagggcga	aaaaccgtct	atcagggcga	4380
	tgcccacta	cgtgaaccat	caccctaata	aagttttttg	gggtcgaggt	gccgtaaagc	4440
50	actaaatcgg	aaccctaaag	ggagcccccg	atttagagct	tgacggggaa	agccggcgaa	4500
	cgtggcgaga	aaggaaggga	agaaagcgaa	aggagcgggc	gctagggcgc	tggcaagtgt	4560
	agcggtcacg	ctgcgcgtaa	ccaccacacc	cgccgcgctt	aatgcgcgcg	tacagggcgc	4620
	gtcccatctg	ccattcaggc	tgcgcaactg	ttgggaaggg	cgatcgggtg	gggcctcttc	4680
	gctattacgc	cagctggcga	aagggggatg	tgctgcaagg	cgattaagtt	gggtaacgcc	4740
55	agggttttcc	cagtcacgac	gttgtaaaac	gacggccagt	gaattgtaat	acgactcact	4800
	atagggcgaa	ttggagctcc	accgcgggtg	c			4831

<210> 100

<211> 20

60 <212> DNA

<213> Artificial Sequence

<220>

65 <223> Description of Artificial Sequence: primer
C31-screen 1

<400> 100

gcgtgagatc aagacgcaca 20

5 <210> 101
 <211> 21
 <212> DNA
 <213> Artificial Sequence

10 <220>
 <223> Description of Artificial Sequence: primer
 C31-screen 2

15 <400> 101
 gcagcggtaa gattccttga t 21

20 <210> 102
 <211> 20
 <212> DNA
 <213> Artificial Sequence

25 <220>
 <223> Description of Artificial Sequence: primer
 beta-Gal 3

30 <400> 102
 atcctctgca tggtcaggtc 20

35 <210> 103
 <211> 18
 <212> DNA
 <213> Artificial Sequence

40 <220>
 <223> Description of Artificial Sequence: primer
 beta-Gal 4

45 <400> 103
 cgtggcctga ttcattcc 18

50 <210> 104
 <211> 5878
 <212> DNA
 <213> Artificial Sequence

55 <220>
 <223> Description of Artificial Sequence: vector
 pCAGGS-Cre-pA

60 <400> 104
 cgccgcgtgc ggccgcgct gcccgccgc tgtgagcgt gccggcgccg ccgccccgctt 60
 tgtgcgctcc gcgtgtgcgc gaggggagcg cggccggggg cggcgccccg cggcgccggg 120
 gggctgcgag gggaacaaag gctgcgtgcg ggggtgtgtc gtgggggggt gagcaggggg 180
 tgtgggcgcg gcggtcgggc tgtaaccccc cctgcaccc cctccccga gttgctgagc 240
 acggccccgc ttcgggtgcg gggtccgtg cggggcggtg cgcggggctc gccgtgccgg 300
 gcgggggggt gcggcaggtg ggggtgcgg gcggggcggg gccgcctcgg gccggggagg 360
 gctcggggga ggggcgcggc ggccccggag cgcggcgggc tgcgaggcg cggcgagccg 420
 cagccattgc cttttatggt aatcgtgcga gagggcgccag ggaacttcct tgtcccaaat 480
 ctggcggagc cgaaatctgg gaggcgccgc cgcacccccct ctacgggcg cgggcgaagc 540
 ggtgcggcgc cggcaggaag gaaatgggcg gggagggcct tcgtgcgtcg ccgcgccgcc 600
 gtccccctct ccatctccag cctcggggct gccgcagggg gacggctgcc ttccgggggg 660
 acggggcagg gcgggggttcg gcttctggcg tgtgaccggc ggctctagta agcgttgggg 720
 tgagtactcc ctctcaaaag cgggcacgac ttctgcgcta agattgtcag tttccaaaaa 780
 cgaggaggat ttgatattca cctggccgcg ggtgatgcct ttgagggtgg ccgcgtccat 840
 ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900

	gccatacact	tgagtgcacat	tgacatccac	tttgcccttc	tctccacagg	tgctccactcc	960
	cagggcgggcc	tcgaccatgc	ccaagaagaa	gagggaaggtg	ttccatttac	tgaccgtaca	1020
	ccaaaatttg	cctgcattac	cggtcgatgc	aacgagtgat	gaggttcgca	agaacctgat	1080
5	ggacatgttc	agggatcgcc	agggcgttttc	tgagcatacc	tggaatatgc	ttctgtccgt	1140
	ttgccgggtcg	tgggcggcat	ggtgcaagtt	gaataaccgg	aaatgggttc	ccgcagaacc	1200
	tgaagatgtt	cgcgattatc	ttctatatct	tcaggcgcg	ggtctggcag	taaaaaactat	1260
	ccagcaacat	ttggggccagc	taaacatgct	tcacgtcgcg	tcgggggtgc	cacgaccaag	1320
	tgacagcaat	gctgtttcac	tggttatgcg	gcggtaccga	aaagaaaacg	ttgatgccgg	1380
10	tgaacgtgca	aaacaggctc	tagcgttcga	acgcactgat	ttcgaccagg	ttcgttcact	1440
	catggaaaaat	agcgatcgct	gccaggatat	acgtaatctg	gcattttctgg	ggattgctta	1500
	taacaccctg	ttacgtatag	ccgaaattgc	caggatcagg	gttaaagata	tctcacgtac	1560
	tgacgggtggg	agaatgttaa	tccatatttg	cagaacgaaa	acgctggta	gcaccgcagg	1620
	tgtagagaag	gcacttagcc	tgggggtaac	taaactggtc	gagcgatgga	tttccgtctc	1680
	tggtgttagct	gatgtccga	ataactacct	gttttgccgg	gtcagaaaaa	atggtgttgc	1740
15	cgcgccatct	gccaccagcc	agctatcaac	tcgcgcctcg	gaagggtatt	ttgaagcaac	1800
	tcacogattg	atttacggcg	ctaaggatga	ctctggtcag	agatacctgg	cctggtctgg	1860
	acacagtgcc	cgtgtcggag	ccgcgcgaga	tatggcccg	gctggagttt	caataccgga	1920
	gatcatgcaa	gctggtggct	ggaccaatgt	aaatattgtc	atgaactata	tccgtaacct	1980
	ggatagttaa	acaggggcaa	tggtgcgcct	gctggaagat	ggcgattagc	cattaacgcg	2040
20	taaatgattg	cagatccact	agttctaggg	ccgcgtcgac	ctcgagatcc	aggcgcggat	2100
	caataaaaaga	tcattatttt	caatagatct	gtgtgttgg	tttttgtgtg	ccttggggga	2160
	gggggaggcc	agaatgaggc	gcggccaagg	gggaggggga	ggccagaatg	accttggggg	2220
	agggggaggc	cagaatgacc	ttgggggagg	gggaggccag	aatgaggcgc	gcccccggt	2280
	accgagctcg	aattcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accttggcgt	2340
25	taccocactt	aatcgccctg	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	2400
	ggcccgacc	gatcgccctt	cccaacagtt	cgccagcctg	aatggcgaat	ggcgctgat	2460
	gcggtatattt	ctccttaagc	atctgtgogg	tatttcacac	cgcatatggt	gcactctcag	2520
	tacaatctgc	tctgatgcgg	catagttaag	ccagcccgca	caccgcgcaa	caccgcctga	2580
	cgcgccctga	cgggcttgtc	tgctcccgcc	atccgcttac	agacaaagctg	tgaccgtctc	2640
30	cgggagctgc	atgtgtcaga	ggttttcacc	gtcatcaccg	aaacgcgcga	gacgaaggg	2700
	cctogtgata	cgctattttt	tataggttaa	tgcatgata	ataatgggtt	cttagacgct	2760
	aggtggcact	tttcggggaa	atgtgcggcg	aaacctatt	tgtttatattt	tctaaataca	2820
	ttcaaatatg	tatccgctca	tgagacaata	acctgataa	atgcttcaat	aatattgaaa	2880
	aaggaagagt	atgagtattc	aacatttccg	tgtcgccctt	attccctttt	ttgcggcatt	2940
35	ttgcttccct	gtttttgtct	accagaaaac	gctgggtgaa	gtaaaagatg	ctgaagatca	3000
	gttgggtgca	cgagtgggtt	acatcgaaat	ggatctcaac	agcggtgaaga	tccttgagag	3060
	ttttcgcccc	gaagaacgtt	ttccaatgat	gagcacittt	aaagtctctg	tatgtggcgc	3120
	ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcggt	cgccgcatac	actattctca	3180
	gaatgaactg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcattgacagt	3240
40	aagagaatta	tgcatgtctg	ccataaccat	gagtgataac	actgocggcca	acttacttct	3300
	gacaacgatc	ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	3360
	aactgcgctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	3420
	caccacgatg	cctgtagcaa	tggaacaac	gttgcgcaaa	ctattaactg	gcgaactact	3480
	tactctagct	tcccggcaac	aattaataga	ctggatggag	gcggataaag	ttgcaggacc	3540
45	acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccgggtga	3600
	gcgtgggtct	cgcggtatca	ttgcagcaat	ggggccagat	ggtaaagcct	cccgatcgt	3660
	agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	3720
	gataggtgcc	tcactgatta	agcattggta	actgtcagac	caagtttact	catatatact	3780
	ttagattgat	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	tcctttttga	3840
50	taatctcatg	accaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	cagaccccg	3900
	agaaaagatc	aaaggatctt	cttgagatcc	ttttttctg	cgcgtaatct	gctgcttgca	3960
	aacaaaaaaa	ccaccgctac	cagcgggtgt	ttgtttgccc	gatcaagagc	taccaactct	4020
	ttttccgaag	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	ttctagtgt	4080
55	gcogtagtta	ggccaccact	tcaagaactc	tgtagcaccc	cctacatacc	tcgctctgct	4140
	aatccctgta	ccagtggctg	ctgccagtgg	cgataagtcg	tgtcttaccg	ggttggactc	4200
	aagacgatag	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggt	cgtgcacaca	4260
	gccagccttg	gagcgaacga	cctacaccca	actgagatac	ctacagcgtg	agctatgaga	4320
	aagcgcacag	cttcccgaag	ggagaaaggc	ggacaggat	ccggttaagc	gcagggtcgg	4380
	aacaggagag	cgcacgaggg	agcttccagg	gggaaacgcc	tggtatcttt	atagtcctgt	4440
60	cgggttttcg	cacctctgac	ttgagcgtcg	atttttgtga	tgctcgtcag	ggggcgagg	4500
	cctatggaaa	aacgccagca	acgcggcctt	tttacgggtc	ctggcctttt	gctggccttt	4560
	tgctcacatg	ttctttctcg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	4620
	tgagtgagct	gataccgctc	gccgcagccg	aacgaccgag	cgcagcaggt	cagtgagcga	4680
	ggaagcggaa	gagcgcacca	tacgcaaac	gcctctcccc	gcgcgttggc	cgattcatta	4740
65	atgcagctgg	cacgacaggt	ttcccgactg	ggaagcggcg	agtgagcgca	acgcaattaa	4800
	tgtagattag	ctcactcatt	aggcaccoca	gttattacac	tttatgcttc	cggctcgtat	4860
	gttgtgtgga	attgtgagcg	gataacaatt	tcacacagga	aacagctatg	accatgatta	4920

86

	cgccaagcta	gccccgggcta	gcttgcacatgc	ctgcagggttt	tgcacattga	ttattgacta	4980
	gttattaata	gtaatcaatt	acgggggtcat	tagttcatag	cccatatatg	gagttccgcg	5040
	ttacataact	tacggtaaat	ggccccgcctg	gctgaccgcc	caacgacccc	cgccattga	5100
5	cgtcaataat	gacgtatgtt	cccatagtaa	cgccaatagg	gactttccat	tgacgtcaat	5160
	gggtggacta	tttacggtaa	actgcccact	tggcagtaca	tcaagtgtat	catatgcca	5220
	gtacgcccc	tattgacgtc	aatgacggta	aatggccgcg	ctggcattat	gcccagtaca	5280
	tgacctatg	ggactttcct	acttggcagt	acatctacgt	attagtcac	gctattacca	5340
	tgggtcgagg	tgagccccc	gttctgcttc	actctcccca	tctccccccc	ctccccacc	5400
	ccaattttgt	atattttat	tttttaatta	ttttgtgcag	cgatgggggc	gggggggggg	5460
10	ggggcgcgcg	ccaggcgggg	cgggcggggg	cgagggggcg	ggcgggggcg	ggcgagagag	5520
	tgcgcgcgca	gccaatcaga	gcggcgcgct	ccgaaagtgt	ccttttatgg	cgaggcgggc	5580
	gcggcgggcg	ccctataaaa	agcgaagcgc	gcggcgggcg	ggagtgcctg	cgttgccttc	5640
	gccccgtgcc	cgcctccgcg	cgcctccgcg	cgcctccgcg	cggctctgac	tgaccgcgtt	5700
	actcccacag	gtgagcgggc	gggacggggc	ttctcctcgc	ggctgttaatt	agcgccttgt	5760
15	ttaatgacgg	ctcgtttcct	ttctgtggct	gcgtgaaagc	cttaaagggc	tccgggaggg	5820
	ccctttgtgc	ggggggggagc	ggctcggggg	gtgcgtgcgt	gtgtgtgtgc	gtggggag	5878

<210> 105

20 <211> 6641

<212> DNA

<213> Artificial Sequence

<220>

25 <223> Description of Artificial Sequence: vector
pCAGGSC31CNLS-pA

<400> 105

30	attgattatt	gactagttat	taatagtaat	caattacggg	gtcattagtt	catagcccat	60
	atatggagtt	ccgcgttaca	taacttaacg	taaatggccc	gcctggctga	ccgcccacag	120
	accccgcgcc	attgacgtca	ataatgacgt	atggttcccat	agtaacgcca	atagggactt	180
	tccattgacg	tcaatgggtg	gactattttac	ggtaaaactgc	ccacttggca	gtacatcaag	240
	tgtatcatat	gccaagtaac	ccccctattg	acgtcaatga	cggtaaatgg	ccgccttggc	300
	attatgcccc	gtacatgacc	ttatgggact	ttcctacttg	gcagtacatc	tacgtattag	360
35	tcacgcctat	taccatgggt	cgagggtgagc	cccacgttct	gcttcactct	ccccatctcc	420
	ccccctccc	cacccccaat	tttgtattta	tttatttttt	aattattttg	tgacgcgatg	480
	ggggcggggg	gggggggggc	gcgcgcgcag	cggggcgggg	cggggcgagg	ggcgggcggg	540
	ggcgaggcgg	agaggtgagg	cggcagccaa	tcagagcggc	gcgctccgaa	agtttccctt	600
	tatggcgagg	cggcggcggc	ggcgcccta	taaaaagcga	agcgcgcggc	ggcggggagt	660
40	cgctgcgttg	ccttcgcccc	gtgccccgct	ccgcgcgcgc	tcgcgcgcgc	cgccccggct	720
	ctgactgacc	gcgttaactc	cacaggtgag	cgggcgggac	ggcccttctc	ctccgggctg	780
	taattagcgc	ttgggttaat	gacggctcgt	ctgttttctg	tggctgcgtg	aaagccttaa	840
	agggctccgg	gaggcccttt	tgctgcgggg	ggagcggctc	gggggggtgc	tgctgtgtgt	900
	tgtgcgtggg	gagcgccggc	tgccgcccgc	gctgcccggc	ggctgtgagc	gctgcggggc	960
45	cggcgcgggg	ccttctgtgc	tcgcgctgtg	cgcgagggga	gcgcggccgg	ggcggtgccc	1020
	ccgcggtgcg	ggggggctgc	gaggggaaca	aaggctgcgt	gcgggggtgt	tgctgtgggg	1080
	ggtgagcagg	gggtgtgggc	gcgcggtcgc	ggctgtaacc	ccccctgca	ccccctccc	1140
	cgagttgctg	agcctccggc	ggcttcgggt	gcgggggtcc	gtgcggggcg	tgccggggcg	1200
	ctgcgcgtgc	cgggcggggg	gtggcgcgag	tggggggtgc	cgggcggggc	ggggcgccct	1260
50	cgggccgggg	agggctcggg	ggaggggcgc	ggcgcccccg	gagcgccggc	ggctgtcgag	1320
	gcgcggcgag	ccgcagccat	tgccctttat	ggtaatcgtg	cgagagggcg	cagggaactc	1380
	ccttctgccc	aatctggcgg	agccgaaatc	tgggagggcg	cgccgcaccc	cctctagcgg	1440
	gcgcgggcga	agcgggtcgg	cgccggcagg	aaggaaatgg	gcggggaggg	ccttcgtgcg	1500
	tcgcgcgcgc	gccgtccctt	tctccatctc	cagcctcggg	gctgcgcgag	ggggacggct	1560
55	gccttcgggg	gggacggggc	agggcggggt	tcggcttctg	gcgtgtgacc	ggcggtctta	1620
	gtaagcgttg	gggtgagtac	tcctctcaaa	aagcgggcat	gacttctgcg	ctaagattgt	1680
	cagtttccaa	aaacgaggag	gatttgatat	tcacctggcc	cgcggtgatg	cctttgaggg	1740
	tggccgcgtc	catctgggtc	gaaaagacaa	tctttttgtt	gtcaagcttg	aggtgtggca	1800
	ggcttgagat	ctggccatac	acttgagtga	cattagacac	cactttgcct	ttctctccac	1860
60	aggtgtccac	tcccaggcgc	gcgcggcgat	atgacacagc	gggttggtgac	cggggtggac	1920
	acgtacgcgg	gtgcttacga	ccgtcagtcg	cgcgagcgcg	agaattcgag	cgcagcaagc	1980
	ccagcgacac	agcgtagcgc	caacgaagac	aaggcgcccg	accttcagcg	cgaagtgcag	2040
	cgcgacgggg	gccggttcag	gttcgtcggg	catttcagcg	aagcgccggg	cacgtcggcg	2100
	ttcgggacgg	cggagcgccc	ggagttcgaa	cgcactcctga	acgaatgccg	cgccggggcg	2160
65	ctcaacatga	tcattgtcta	tgacgtgtgc	cgtctctcgc	gcctgaaggt	catggacgcg	2220
	attccgattg	tctcggtaatt	gtcgcgctgc	ggcgtgacga	ttgtttccac	tcagggaagg	2280
	gtcttccggc	agggaaacgt	catggacctg	attcaacctga	ttatgcggct	cgacgcgtcg	2340

	cacaaagaat	cttcgctgaa	gtcggcggaag	attctcgaca	cgaagaacct	tcagcgcgaa	2400
	ttgggcgggt	acgtcggcgg	gaaggcgctt	tacggcttcg	agcttggttc	ggagacgaag	2460
	gagatcacgc	gcaacggcgg	aatgggtcaat	gtcgtcatca	acaagcttgc	gcactcgacc	2520
	actcccctta	ccggaccctt	cgagttcgag	cccgaacgtaa	tcgggtgggtg	gtggcgtgag	2580
5	atcaagacgc	acaaacacct	tcccttcaag	ccgggcagtc	aagccgccat	tcacccgggc	2640
	agcatcacgg	ggctttgtaa	gcgcattggac	gctgacgcgg	tgccgacccg	gggcgagacg	2700
	attgggaaga	agaccgcttc	aagcgcctgg	gacccggcaa	ccgttatgcg	aatccttcgg	2760
	gacccgcgta	ttgcgggctt	cgccgctgag	gtgatctaca	agaagaagcc	ggacggcagc	2820
	ccgaccacga	agattgaggg	ttaccgcatt	cagcgcgacc	cgatcacgct	ccggccgggtc	2880
10	gagcttgatt	gcggaccgat	catcgagccc	gctgagtggg	atgagcttca	ggcgtgggtg	2940
	gacggcaggg	ggcgcgccaa	ggggctttcc	cgggggcaag	ccattctgtc	cgccatggac	3000
	aagctgtact	gcgagtgtgg	cgccgtcatg	acttcgaagc	gcggggaaga	atcgatcaag	3060
	gactcttacc	gctgcgctcg	ccggaaggtg	gtcgacccgt	ccgcacctgg	gcagcacgaa	3120
	ggcagctgca	agctcagcat	ggcggcagtc	gacaaagttc	ttgcggaaac	catcttcaac	3180
15	aagatcaggc	acgcgcgaag	cgacgaagag	acgtttggcg	ttctgtggga	agccgcccga	3240
	cgcttcggca	agctcactga	ggcgccctgag	aagagcggcg	aacgggcgaa	ccttggttgcg	3300
	gagcgcgcgg	acgccttgaa	cgcccttgaa	gagctgtacg	aagaccgcgc	ggcaggcgcg	3360
	tacgacggac	ccgtttggcag	gaagcaactt	cggaagcaac	aggcagcgct	gacgctccgg	3420
	cagcaggggg	cggaagagcg	gcttgccgaa	cctgaagccg	ccgaagcccc	gaagcttccc	3480
20	cttgaccaat	ggttccccga	agacgcgcag	gctgacccga	ccggccctaa	gtcgtgggtg	3540
	gggcgcgcgt	cagtagacga	caagcgcgtg	ttcgtcgggc	tcttcgtaga	caagatcggt	3600
	gtcacgaagt	cgactacggg	cagggggcag	ggaacgcccc	tcgagaagcg	cgcttcgatc	3660
	acgtgggcga	agccgcgcag	cgacgacgac	gaagacgacg	cccaggacgg	cacggaagac	3720
	gtagcggcgc	ctaagaagaa	gaggaaggtt	tagactctcg	agatccaggc	gcggatcaat	3780
25	aaaagatcat	tatttttcaat	agatctgtgt	tttgggtttt	tgtgtgcctt	gggggagggg	3840
	gaggccagaa	tgaggcgcgg	ccaaggggga	gggggagggc	agaatgacct	tgggggaggg	3900
	ggaggccaga	atgaccttgg	gggaggggga	ggccagaatg	aggcgcgccc	ccgggtaccg	3960
	agctcgaatt	cactggccgt	cgttttacaa	cgctgtgact	gggaaaaccc	tggcgttacc	4020
	caacttaatc	gccttgccag	acatccccct	ttcgccagct	ggcgtaatat	cgaagaggcc	4080
30	cgcacccgat	gcccttcccc	acagttgcgc	agcctgaatg	gcgaatggcg	cctgatgcgg	4140
	tatttttctc	ttacgcctct	gtgcgggtat	tcacaccgca	tatggtgcac	tctcagtaac	4200
	atctgctctg	atgcgcgata	gttaagccag	ccccgacacc	cgccaacacc	cgctgacggc	4260
	ccctgacggg	cctgtctgtt	cccgccatcc	gcttacagac	aagctgtgac	cgtctccggg	4320
	agctgcatgt	gtcagaggtt	ttcaccgtca	tcaccgaaac	gcgcgagacg	aaagggcctc	4380
35	gtgatacgcc	tattttttata	ggttaatgtc	atgataataa	tggtttctta	gacgtcaggt	4440
	ggcacttttc	ggggaaatgt	gcgcgggaac	cctattttgt	tatttttcta	aatacattca	4500
	aatatgtatc	acaataaccc	gtataaatgc	gataaaatgc	ttcaataata	ttgaaaaagg	4560
	aagagtatga	gtattcaaca	tttccgtgtc	goccttattc	ccttttttgc	ggcatttttgc	4620
	cttctctgtt	ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	4680
40	ggtgcacgag	tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	4740
	cgccccgaag	aacgttttcc	aatgatgagc	acttttaaa	ttctgctatg	tggcgcggtg	4800
	ttatcccgta	ttgacgcggg	gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	4860
	gacttggttg	agtcacagaa	agtcactcta	gataaacactg	cggatggcat	gacagtaaga	4920
45	gaattatgca	gtgctgccat	aaccatgagt	tttttgacac	cgcccaactt	acttctgaca	4980
	acgatcggag	gaccgaagga	gctaaccgct	tttttgacac	acatggggga	tcagttaact	5040
	cgcccttgatc	gttgggaacc	ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	5100
	acgatgcttg	tagcaatggc	aacaacgctg	cgcaaaactat	taactggcga	actacttact	5160
	ctagcttccc	ggcaacaatt	aatgagcgcg	attggagcgg	ataaagttgc	aggaccactt	5220
	ctgcgctcgg	cccttcgggc	tgggtggttt	atgtgctgata	aatctggagc	cggtgacgct	5280
50	gggtctcgcg	gtatcattgc	agcactgggg	ccagatggta	agccctcccg	tatcgtagtt	5340
	atctacacga	cggggagtc	ggcaactatg	gatgaacgaa	atagacagat	cgctgagata	5400
	ggtgcctcac	tgattaaagc	ttggtaactg	tcagaccaag	tttactcata	tatactttag	5460
	attgatttaa	aacttcaatt	ttaatttaaa	aggatctagg	tgaagatcct	ttttgataat	5520
	ctcatgacca	aaatccctta	acgtgagttt	tcgttccact	gagcgtcaga	ccccgtagaa	5580
55	aagatcaaag	gatcttcttg	agatcccttt	tttctgcgcg	taatctgctg	ccttgcaaca	5640
	aaaaaaccac	cgctaaccag	ggtggtttgt	ttgcccggatc	aagagctacc	aactcttttt	5700
	ccgaaggtaa	ctggttcag	cagagcgcg	ataccaaata	ctgtccttct	agtgtagcgg	5760
	tagttaggcc	accaottcaa	gaactctgta	gcaccgccta	catacctcgc	tctgctaate	5820
	ctgtttaccg	tggctgctgc	cagtggcgat	aagtcgtgtc	ttaccgggtt	ggactcaaga	5880
60	cgatagttac	cggataaggc	gcagcggtcg	ggctgaacgg	gggggttcgtg	cacacagcc	5940
	agcttggagc	gaacgacct	caccgaactg	agatacctac	agcgtgagct	atgagaaagc	6000
	gccacgcttc	ccgaagggag	aaaggcggac	aggtatccgg	taagcggcag	ggtcgggaaca	6060
	ggagagcgca	cgagggagct	tccaggggga	aacgcctggt	atctttatag	tcctgtcggg	6120
	tttcgccacc	tctgacttga	gcgtcgattt	ttgtgatgct	cgtcaggggg	gcggagccta	6180
65	tggaaaaacg	ccagcaacgc	ggccttttta	cgttccctgg	ccttttgctg	gccttttgct	6240
	cacatgttct	ttccctcggt	atcccttgat	ctgtggata	accgtattac	cgcccttgag	6300
	tgagctgata	ccgctcgccg	cagccgaacg	accgagcgca	gcgagtcagt	gagcggaggaa	6360

5 ggcgaagagc gcccaatacg caaaccgcct ctcccgcgc gttggccgat tcattaatgc 6420
 agctggcaag acaggtttcc cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg 6480
 agttagctca ctcataggc accccaggct ttacacttta tgcttcgggc tcgtatgttg 6540
 tgtggaattg tgagcggata acaatttcac acaggaaaca gctatgacca tgattacgcc 6600
 aagctagccc gggctagctt gcatgcctgc aggttttcga c 6641

10 <210> 106
 <211> 11784
 <212> DNA
 <213> Artificial Sequence

15 <220>
 <223> Description of Artificial Sequence: modified
 ROSA26 locus

20 <400> 106
 ggcaggccct ccgagcgtgg tggagccgtt ctgtgagaca gccgggtacg agtcgtgacg 60
 ctggaagggg caagcgggtg gtgggcagga atgcggtcgg cctgcagca accggagggg 120
 gagggagaag ggagcggaaa agtctccacc ggacgcggcc atggctcggg gggggggggg 180
 cagcggagga gcgcttccgg ccgacgtctc gtcgctgatt ggcttctttt cctcccggcg 240
 tgtgtgaaaa cacaaatggc gtgttttggg tggcgtaagg cgctgtcag ttaacggcag 300
 ccggagtgcg cagccggccg cagcctcgct ctgcccactg ggtggggcgg gaggtagggtg 360
 ggggtgaggcg agctggacgt gcgggcgcgg tcggcctctg gcggggcggg ggaggggagg 420
 25 gagggtcagc gaaagtagct cgcgcgcgag cggcgcccca cctcccctt cctctggggg 480
 agtcgtttta cccggcccg cccgggacct gtctgtgat tggctctcgg ggcccagaaa 540
 actggccctt gccattggct cgtgttcgtg caagttagat ccatccgcgg gccagcgggg 600
 gcggcgagga ggcgtccca ggttcgggcc ctcccctcgg ccccgcgccg cagagtctgg 660
 ccgcgcgccc ctgcgcaacg tggcaggaag cgcgcgctgg gggcggggac gggcagtagg 720
 30 gctgagcggc tgcggggcgg gtgcaagcac gtttcggact tgagttgcct caagaggggc 780
 gtgctgagcc agacctccat cgcgcactcc ggggagtggg ggggaaggagc gagggctcag 840
 ttgggtgtgt ttggaggcag gaagcacttg ctctccaaa gtcgctctga gttgttatca 900
 gtaaggagc tgcagtggag taggcgggga gaaggccgca ccttctccg gaggggggag 960
 gggagtgttg caataccttt ctgggagttc tctgctgcct cctggcttct gaggaccggc 1020
 35 ctgggccttg gagaatccct tcccctctt cctcgtgat ctgcaactcc agtctttcgc 1080
 ctaggtaacc gatatccctg caggggtgac ctgcacgtct agggcgagc agtccagggt 1140
 ttctgtatg atgtcactat tatctgttcc ttttttttcc ctcagctcgc ggttgaggac 1200
 aaactcttgc cggctcttcc agtactcctg caggtgactg actgagtcga cgacactgca 1260
 40 gagacctact tcaactaaca ccggtacagt tctgtgacca gatgggtgag gtggagtacg 1320
 cgcccgggga gcccaagggc acccctggc acccgacccg cggcttogag accgtcacga 1380
 ataacttcgt atagcataca ttatacgaag ttataagctc gatgaattct accgggtagg 1440
 ggaggcgctt ttcccaaggc agtctggagc atgcgcttta gcagcccgc tggcactttg 1500
 cgctacacaa ttggcctctg gectcgaca cattccacat ccacoggtag cgccaaccgg 1560
 45 ctccgttctt tgggtggccc ttgcgcgccac cttctactcc tcccctagtc aggaagttcc 1620
 ccccgccccc gcagctcgcg tcgtgcagga cgtgacaaat ggaagtagca cgtctcacta 1680
 gtctcgtgca gatggacagc accgctgagc aatggaagcg ggtaggcctt tggggcagcg 1740
 gccaatagca gctttgctcc ttgcgtttct gggctcagag gctgggaagg ggtgggtccg 1800
 ggggcgggct caggggcggg ctcaaggcgc gggcgggcgc gaaggctctc ccgagcccg 1860
 50 gcattctcgc acgcttcaaa agcgcacgtc tgcgcgctg ttctcctctt cctcatctcc 1920
 gggcctttcg acgatccagc gccaccatg aaaaagcctg aactcacgc gacgtctgtc 1980
 gagaagtttc tgatcgaaaa gttcgacagc gtctccgacc tgatgcagct ctggaggggc 2040
 gaagaatctc gtgctttcag cttecatgta ggaggcggtg gatatgtcct gcgggtaaat 2100
 agctgcgcgg atggtttcta caaagatcgt tatgtttatc ggcactttgc atcgcccgcg 2160
 55 ctcccgattc cggaaagtgc tgacattggg gaattcagcg agagcctgac ctattgcac 2220
 tcccgcggtg cacagggtgt caggttgcaa gacctgctg aaaccgaact gccgcgtgtt 2280
 ctgcagccgg tcgcggaggc catggatgcg atcgtgcgg ccgatcttag ccagacgagc 2340
 ggggttcggc cattcggaac gcaaggaaat ggtcaataca ctacatggcg tgatttcata 2400
 tgcgcgattg ctgatcccca tgtgtatcac tggcaaaact tgatggacga caccgtcagt 2460
 60 gcgtccgtcg cgcaggctct cgatgagctg atcctttggg ccgaggactg ccccgagtc 2520
 cggcacctcg tgcacgcgga tttcggtccc aacaatgtcc tgacggacaa tggccgcata 2580
 acagcgggtc ttgactggag cgaggcgatg ttoggggatt cccaatacga ggtcgccaac 2640
 atcttcttct ggaggccgtg gttggcttgt atggagcagc agacgcgcta cttcgagcgg 2700
 aggcacccgg agcttgagg atcgccgcgg ctccgggctg atatgctccg cattgggtctt 2760
 65 gaccaactct atcagagctt ggttgaoggc aatttcgatg atgcagcttg ggcgcagggt 2820
 cgatgcgagc caatcgtccg atccggagcc gggactgtcg ggcgtacaca aatgcgccg 2880
 agaagcgagg cgtctggac cgatggtgtg gtagaagtac tggccgatag tggaaaccga 2940
 cgcccagca ctgctccgag ggcaaaggaa tagtcgatgc agaaattgat gatctattaa 3000

	acaataaaga	tgtccactaa	aatggaagtt	tttctgtgca	tactttgtta	agaagggtga	3060
	gaacagagta	cctacatttt	gaatggaagg	attggagcta	cgggggtggg	ggtgggggtgg	3120
	gattagataa	atgootgctc	tttactgaag	gctctttact	attgctttat	gataatgttt	3180
5	catagttgga	tatcataatt	taaacaagca	aaaccaaatt	aagggccagc	tcatctctcc	3240
	cactcatgat	ctatagatct	atagatctct	cgtgggatca	ttgtttttct	cttgattccc	3300
	acttttgtgt	tctaagtact	gtggtttcca	aatgtgtcag	tttcatagcc	tgaagaacga	3360
	gatcagcagc	ctctgtttcca	catacaacttc	attctcagta	ttgtttttgcc	aagttctaat	3420
	tccatcagaa	gcttcagctg	ctcgactaga	ggatcataat	cagccatacc	acattttag	3480
	aggtttttact	tgctttaaaa	aacctccac	acctccccc	gaacctgaaa	cataaaatga	3540
10	atgcaattgt	tggtgttaac	ttgtttattg	cagcttataa	tggttacaaa	taaagcaata	3600
	gcacacaaa	tttcacaaat	aaagcatttt	tttcaactgca	ttctagttgt	ggtttgtcca	3660
	aactcatcaa	tgtatcttat	catgtctgga	tccgtgtcat	gtcggcgacc	ctacgcccc	3720
	aactgagaga	actcaaaggt	taccccagtt	ggggcactac	tcccgaatac	cgcttctgga	3780
	tccataactt	cgtatagcat	acattatacg	aagttatacc	ggggccaccat	ggtcgcgagt	3840
15	agcttggcac	tgcccgctgt	tttacaacgt	cgtgactggg	aaaacctgtg	cgttacccaa	3900
	cttaatcgcc	ttgcagcaca	tccccctttc	gccagctggc	gtaatagcga	agaggcccg	3960
	accgatcgcc	cttcccaaca	gttgccgcagc	ctgaatggcg	aatggcgctt	tgctgtggtt	4020
	ccggcaccag	aagcgggtgc	ggaaagctgg	ctggagtgcg	atcttctctga	ggccgatact	4080
	gtcgtcgtcc	cctcaaaactg	gcagatgcac	ggttacgatg	cgcccatcta	caccaacgta	4140
20	acctatccca	ttacgggtcaa	tccgcggttt	gttcccacgg	agaatccgac	gggttgttac	4200
	tcgctcacat	ttaatgttga	tgaaagctgg	ctacaggaag	gccagacggc	aattattttt	4260
	gatggcggtt	actcggcggt	tcactctgtg	tgcaacgggc	gctgggtcgg	ttacggccag	4320
	gacagtctgt	tgccgtctga	atlttgacctg	agcgcatttt	tacgcgcggg	agaaaaccgc	4380
	ctcgcggtga	tggtgctgcg	ttggagtgc	ggcagttatc	tggaagatca	ggatatgtgg	4440
25	cggatgagcg	gcattttccg	tgacgtctcg	ttgctgcata	aaccgactac	acaaactcagc	4500
	gatttccatg	tttgcaactcg	ctttaatgat	gatttcagcc	gcgctgtact	ggaggctgaa	4560
	gttcagatgt	gcggcgagtt	gcgtgactac	ctacgggtaa	cagtttcttt	atggcagggg	4620
	gaaacgcagg	tcgccagcgg	caccgcgcct	ttcggcgggtg	aaattatcga	tgagcgtggg	4680
	ggttatgccg	atcgcgtcac	actacgtctg	aacgtcgaaa	acccgaaaact	gtggagcgcc	4740
30	gaaatcccca	atctctatcg	tgccgtgggt	gaactgcaca	ccgccgacgg	cacgctgatt	4800
	gaagcagaag	cctgcgatgt	cggtttccgc	gaggtgcgga	ttgaaaatgg	tctgctgctg	4860
	ctgaacggca	agccgttgct	gattcgaggg	gttaaccgtc	acgagcatca	tctctgtcat	4920
	ggtcagggtc	tggtgagcga	gacgatgggtg	caggatatcc	tgctgatgaa	gcagaacaac	4980
	tttaacgcgc	tgccgtgttc	gcattatccg	aaccatccgc	tggtgtacac	gctgtgcgac	5040
35	cgtacggcc	tgtatgtggt	ggatgaagcc	aattattgaa	cccacggcat	ggtgccaatg	5100
	aatcgtctga	ccgatgatcc	gcgctggcta	ccggcgatga	gcgaacgcgt	aacgcgaatg	5160
	gtgcagcgcg	atcgtaatca	cccagtgctg	atcatctggt	cgtggggaa	tgaatcaggc	5220
	cacggcgcta	atcacgacgc	gctgtatcgc	tggtacaaat	ctgtcgatcc	ttcccgcgcg	5280
	gtgcagtatg	aaggcggcgg	agccgacacc	acggccaccg	atattatttg	cccgatgtac	5340
40	gcgcgcgtgg	atgaagacca	gcccttcccg	gctgtgcgga	aatggtccat	caaaaaatgg	5400
	ctttcgtctac	ctggagagac	gcgcgcgctg	atccttttgcg	aatacgccca	cgcgatgggt	5460
	aacagttctt	gcggtttccg	taaataactgg	caggcgtttc	gtcagtatcc	ccgtttacag	5520
	ggcggcttcg	tctgggactg	ggtggatcag	tcgctgatta	aatatgatga	aaacggcaac	5580
45	ccgtggtcgg	cttacggcgg	tgatttttggc	gatacgcgga	acgatcgcca	gttctgtatg	5640
	aacgggtctg	tctttgccga	ccgcacgcgc	catccagcgc	tgacggaagc	aaaacaccag	5700
	cagcagtttt	tccagttccg	tttatccggg	caaaccatcg	aagtgaccag	cgaataacctg	5760
	ttccgtcata	gcgataacga	gctcctgcac	tggtatgggtg	cgctggatgg	taagccgctg	5820
	gcaagcgggtg	agtgccctct	ggatgtcgtc	ccacaaggta	aacagttgat	tgaactgcct	5880
	gaactacccg	agccggagag	cgccgggcaa	ctctggctca	cagtacgcgt	agtgcacccg	5940
50	aacgcgaccc	catggtcaga	agccggggcac	atcagcgcct	ggcagcagtg	gcgtctggcg	6000
	gaaaacctca	gtgtgacgct	ccccgcgcgc	tcccacgcca	tcccgcatct	gaccaccagc	6060
	gaaatggatt	tttgcatcga	gctgggtaat	aagcgttggc	aatttaaccg	ccagtcaggc	6120
	tttctttcac	agatgtggat	tgccgataaa	aaacaactgc	tgacgcgcgt	gcgcgatcag	6180
	ttcaccctgt	caccgtctga	taacgacatt	ggcgtaagt	aagcgaccgc	cattgacct	6240
55	aacgcctggg	tcgaacgctg	gaaggcggcg	ggccattacc	aggccgaagc	agcgttgttg	6300
	cagtgcacgg	cagatacact	tgctgatgcg	gtgctgatta	cgaccgctca	cgctgtggcag	6360
	catcagggga	aaaccttatt	tatcagccgg	aaaacctacc	ggattgatgg	tagtgggtcaa	6420
	atggcgatta	ccgttgatgt	tgaagtggcg	agcgatacac	cgcatccggc	gcggattggc	6480
	ctgaactgcc	agctggcgca	ggtagcagag	cgggtaaaact	ggctcggatt	agggccgcaa	6540
60	gaaaactatc	ccgacgcct	tactgcggcc	tgttttgacc	gctgggatct	gccattgtca	6600
	gacatgtata	ccccgtacgt	cttcccagac	gaaaacgggtc	tgccgtgcgg	gacgcgcgaa	6660
	ttgaattatg	gcccacacca	gtggcgcggc	gacttccagt	tcaacatcag	ccgctacagt	6720
	caacagcaac	tgatggaaac	cagccatcgc	catctgctgc	acgcggaaga	aggcacatgg	6780
	ctgaatatcg	acggtttcca	tatggggatt	ggtggcgacg	actcctggag	cccgctagta	6840
65	tcggcggaat	tccagctgag	cgccggtcgc	taccattacc	agttggctctg	gtgtcaaaaa	6900
	taataataac	cgggcagggg	ggatctttgt	gaaggaaact	tacttctgtg	gtgtgacata	6960
	attggacaaa	ctacctacag	agattttaag	acttaaggta	aatataaaat	ttttaagtgt	7020

	ataatgtggt	aaactactga	ttctaattgt	tttgttattt	tagattccaa	cctatggaac	7080
	tgatgaatgg	gagcagtggg	ggaatgccag	atccagacat	gataagatac	attgatgagt	7140
	ttggacaaac	cacaactaga	atgcagtga	aaaaatgctt	tatttgtgaa	atttgtgatg	7200
5	ctattgcttt	atttgaacc	attataagct	gcaataaaca	agttaacaac	aacaattgca	7260
	ttcattttat	gtttcagggt	cagggggagg	tgtgggaggt	tttttaaagc	aagtaaaacc	7320
	tctacaaatg	tggtatggct	gattatgatc	tgccggccaaa	tcggccggcc	tagggcggcc	7380
	ggtaaccgaa	gttcctatc	tttctagaga	ataggaactt	cggaatagga	acttcaagct	7440
	taagcgctag	cctagaagat	ggcggggagt	cttctgggca	ggcttaaagg	ctaaccctgg	7500
	gtgtgggcgt	tgctctgcag	gggaattgaa	caggtgtaaa	attggaggga	caagacttcc	7560
10	cacagatttt	cggttttgtc	gggaagtttt	ttaatagggg	caaataagga	aaatgggagg	7620
	ataggtagtc	atctgggggt	ttatgcagca	aaactacagg	ttattattgc	ttgtgatccg	7680
	cctcggagta	ttttccatcg	aggtagatta	aagacatgct	cacccgagtt	ttatactctc	7740
	ctgcttgaga	tccttactac	agtatgaaat	tacagtgctg	cgagttagac	tatgtaagca	7800
	gaattttta	cattttttaa	gagccagcta	cttcataatc	atttctcccg	ctccttctgc	7860
15	agccttatca	aaaggatatt	tagaacactc	atttttagccc	catttttcatt	tattatactg	7920
	gcttatccaa	cccctagaca	gagcattggc	attttccctt	tcctgatctt	agaagtctga	7980
	tgactcatga	aaccagacag	attagttaca	tacaccacaa	atcgaggctg	tagctggggc	8040
	ctcaacactg	cagttctttt	ataactcctt	agtacacttt	ttgttgatcc	tttgccttga	8100
	tccttaattt	tcagtgctct	tcacctctcc	cgctcagtggt	gttccacatt	tgggcctatt	8160
20	ctcagtcocg	ggagttttac	aacaatagat	gtattgagaa	tccaacctaa	agcttaactt	8220
	tccaactccc	tgaatgcctc	tctccttttt	ctccatttat	aaactgagct	attaaccatt	8280
	aatgggtcca	ggtggatgtc	tcctcccat	attacctgat	gtatcttaca	tattgccagg	8340
	ctgatatttt	aagacattaa	aaggtatatt	tcattattga	gccacatggt	attgattact	8400
	gcttactaaa	attttgtcat	tgtacacatc	tgtaaaaggt	ggttcccttt	ggaatgcaaa	8460
25	gttcaggtgt	ttgttgtctt	tcctgacctc	aggtcttgtg	agcttgtatt	ttttctattt	8520
	aagcgatgct	ttctcttggg	ctggtctgac	tcattggcatt	ctacacgtta	ttgctggtct	8580
	aaatgtgatt	ttgccaagct	tcctcaggac	ctataatttt	gcttgacttg	tagccaaaca	8640
	caagtaaaat	gattaagcaa	caaatgtatt	tgtgaagctt	ggtttttagg	ttgttgtgtt	8700
	gtgtgtgctt	gtgctctata	ataatactat	ccaggggctg	gagagggtgg	tcggagttca	8760
30	agagcacaga	ctgctcttcc	agaagtctct	agttcaattc	ccagcaacca	catgggtggc	8820
	cacaaccatc	tgtaatggga	tctgatggcc	tctctgggtg	tgtctgaaga	ccacaagctg	8880
	attcacatta	aataaataaa	tcctccttct	tcttcttttt	ttttttttta	aagagaatac	8940
	tgtctccagt	agaatttact	gaagtaatga	aatactttgt	gtttgttcca	atatggtagc	9000
	caataatcaa	attactcttt	aagcactgga	aatgttacca	aggaactaat	ttttatttga	9060
35	agtgtaaactg	tggacagagg	agccataact	gcagacttgt	gggatacaga	agaccaatgc	9120
	agactttaat	gtcttttctc	ttacactaag	caataaagaa	ataaaaaattg	aacttctagt	9180
	atcctatttg	tttaaactgc	tagctttact	taacttttgt	gcttcactca	tacaaagctg	9240
	aaagctaagt	ctgcagccat	tactaaacat	gaaagcaagt	aatgataatt	ttggatttca	9300
	aaaatgtagg	gccagagttt	agccagccag	tgggtggtgct	tgccctttatg	cttttaattcc	9360
40	cagcactctg	gaggcagaga	caggcagatc	tctgagtttg	agcccagcct	ggtctacaca	9420
	tcaagttcta	tctaggatag	ccaggaatac	acacagaaac	cctgttgggg	aggggggctc	9480
	tgagatttca	taaaattata	attgaagcat	tcctaatga	gccactatgg	atgtggctaa	9540
	atcctgtcac	ctttctgatg	agatttgggt	ctgtctctgc	ctgtctctgc	tggttggttg	9600
	gtcttttgac	actgtgggct	ttcttttaag	cctccttcc	gccatgtggt	ctcttgtttg	9660
45	ctactaactt	cccattggct	aaatggcatg	gctttttgct	ttctaagggc	agctgctgag	9720
	atttgcagcc	tgatttccag	ggtgggggtg	ggaaatcttt	caaacactaa	aattgtcctt	9780
	taattttttt	tttaaaaaat	gggttatata	ataaacctca	taaaatagtt	atgaggagtg	9840
	aggtggacta	atattaaatg	agtcctctcc	ctataaaaga	gctatttaagg	cttttctgt	9900
	tatacttaac	ttttttttta	aatgtggtat	ctttagaacc	aagggtctta	gagttttagt	9960
50	atacagaaac	tggtgcatcg	cttaatcaga	ttttctagtt	tcaaatccag	agaatccaaa	10020
	ttcttcacag	ccaaagtcaa	attaagaatt	tctgactttt	aatgttaatt	tgcttactgt	10080
	gaatataaaa	atgatagctt	ttcctgaggc	agggctctcac	tatgtatctc	tgccctgatc	10140
	gcaacaagat	actgtagacta	aagttctgcc	tgcttttgtc	tcctgaatac	taagggttaa	10200
	atgtagtaat	cttcttgtaa	cttgccggtc	agattctttt	ataggggaca	cactaaggga	10260
55	gcttgggtga	tagtttgtaa	aatgtgtttc	aagtgtatga	aacttgaatt	attatccagg	10320
	caacctactt	tttaaaaaaa	aaagccaggc	ctgttagagc	atgcttaagg	gatccctagg	10380
	acttgctgag	cacacaagag	tagttacttg	gcaggctcct	ggtgagagca	tatttcaaaa	10440
	aacaaggcag	acaaccaaga	aactacagtt	aaggttacct	gtctttaaac	catctgcata	10500
	tacacaggga	tattaaaata	ttccaaataa	tatttcatte	aagttttccc	ccatcaaat	10560
60	gggacatgta	tttctccggt	gaataggcag	agttggaaac	taaacaaatg	ttggttttgt	10620
	gatttgtgaa	attgttttca	agtgtatggt	aaagcccatg	agatacagaa	caaagctgct	10680
	atttgcagggt	ctcttgggtt	atactcagaa	gcacttcttt	gggtttccct	gcactatcct	10740
	gatcatgtgc	taggcctacc	ttaggctgat	tgttgttcaa	ataaacctaa	gtttcctgtc	10800
	aggtgatgtc	atatgatttc	atatatcaag	gcaaaacatg	ttatatatgt	taaacatttg	10860
65	tacttaatgt	gaaagttagg	tccttctggg	tttgattttt	aattttcaaa	acctgagcta	10920
	aataagtcac	ctttacattg	ctttacattg	gtggaattgt	ataattgtgg	tttgaggcca	10980
	agactctctg	acctagtaac	cctacctata	gagcaatttg	ctgggtcaca	agtctaggag	11040

91

	tcaagcattt	caccttgaag	ttgagacggt	ttggttagtgt	atactagttt	atatgttgga	11100
	ggacatgttt	atocagaaga	tattcaggac	tatttttgac	tgggctaagg	aattgattct	11160
	gattagcact	gttagtgagc	attgagtggc	ctttaggcct	gaattggagt	cacttgtata	11220
5	tctcaaataa	tgctggcctt	ttttaaaaag	cccttggtct	ttatcaccct	gttttctaca	11280
	taatttttgt	tcaaagaaat	acttgtttgg	atctcctttt	gacaacaata	gcatgttttc	11340
	aagccatatt	ttttttcctt	tttttttttt	tttttggttt	ttcgagacag	ggtttctctg	11400
	tatagccctg	gctgtcctgg	aactcacttt	gtagaccagg	ctggcctcga	actcagaaat	11460
	ccgcctgcct	ctgcctcctg	agtgcgggga	ttaaaggcgt	gcaccaccac	gcctgggctaa	11520
10	gttgatatt	ttgttatata	actataacca	atactaactc	cactgggtgg	atttttaatt	11580
	cagtcagtag	tcttaagtgg	tctttatttg	cccttcatta	aaatctactg	ttcactctaa	11640
	cagaggctgt	tggtactagt	ggcacttaag	caacttcocta	cggatatact	agcagattaa	11700
	gggtcaggga	tagaaactag	tctagcgttt	tgtataccta	ccagctttat	actaccttgt	11760
	tctgatagaa	atatttcagg	acat				11784
15	<210> 107						
	<211> 1458						
	<212> DNA						
	<213> Bacteriophage TP901-1						
20	<220>						
	<221> CDS						
	<222> (1)..(1455)						
25	<400> 107						
	atg act aag aaa gta gca atc tat aca cga gta tcc act act aac caa	48					
	Met Thr Lys Lys Val Ala Ile Tyr Thr Arg Val Ser Thr Thr Asn Gln						
	1 5 10 15						
30	gca gag gaa ggg ttc tca att gat gag caa att gac cgt tta aca aaa	96					
	Ala Glu Glu Gly Phe Ser Ile Asp Glu Gln Ile Asp Arg Leu Thr Lys						
	20 25 30						
35	tat gct gaa gca atg ggg tgg caa gta tct gat act tat act gat gct	144					
	Tyr Ala Glu Ala Met Gly Trp Gln Val Ser Asp Thr Tyr Thr Asp Ala						
	35 40 45						
40	ggt ttt tca ggg gcc aaa ctt gaa cgc cca gca atg caa aga tta atc	192					
	Gly Phe Ser Gly Ala Lys Leu Glu Arg Pro Ala Met Gln Arg Leu Ile						
	50 55 60						
45	aac gat atc gag aat aaa gct ttt gat aca gtt ctt gta tat aag cta	240					
	Asn Asp Ile Glu Asn Lys Ala Phe Asp Thr Val Leu Val Tyr Lys Leu						
	65 70 75 80						
50	gac cgc ctt tca cgt agt gta aga gat act ctt tat ctt gtt aag gat	288					
	Asp Arg Leu Ser Arg Ser Val Arg Asp Thr Leu Tyr Leu Val Lys Asp						
	85 90 95						
55	gtg ttc aca aaa aat aaa ata gac ttt atc tcg ctt aat gaa agt att	336					
	Val Phe Thr Lys Asn Lys Ile Asp Phe Ile Ser Leu Asn Glu Ser Ile						
	100 105 110						
60	gat act tct tct gct atg ggt agc ttg ttt ctc act att ctt tct gca	384					
	Asp Thr Ser Ser Ala Met Gly Ser Leu Phe Leu Thr Ile Leu Ser Ala						
	115 120 125						
65	att aat gag ttt gaa aga gag aat ata aaa gaa cgc atg act atg ggt	432					
	Ile Asn Glu Phe Glu Arg Glu Asn Ile Lys Glu Arg Met Thr Met Gly						
	130 135 140						
70	aaa cta ggg cga gcg aaa tct ggt aag tct atg atg tgg act aag aca	480					
	Lys Leu Gly Arg Ala Lys Ser Gly Lys Ser Met Met Trp Thr Lys Thr						
	145 150 155 160						
75	gct ttt ggg tat tac cac aac aga aag aca ggt ata tta gaa att gtt	528					
	Ala Phe Gly Tyr Tyr His Asn Arg Lys Thr Gly Ile Leu Glu Ile Val						

				165				92		170				175			
5	cct	tta	caa	gct	aca	ata	gtt	gaa	caa	ata	ttc	act	gat	tat	tta	tca	576
	Pro	Leu	Gln	Ala	Thr	Ile	Val	Glu	Gln	Ile	Phe	Thr	Asp	Tyr	Leu	Ser	
				180					185					190			
10	gga	ata	tca	ctt	aca	aaa	tta	aga	gat	aaa	ctc	aat	gaa	tct	gga	cac	624
	Gly	Ile	Ser	Leu	Thr	Lys	Leu	Arg	Asp	Lys	Leu	Asn	Glu	Ser	Gly	His	
			195					200					205				
15	atc	ggt	aaa	gat	ata	ccg	tgg	tct	tat	cgt	acc	cta	aga	caa	aca	ctt	672
	Ile	Gly	Lys	Asp	Ile	Pro	Trp	Ser	Tyr	Arg	Thr		Leu	Arg	Gln	Thr	
		210					215					220					
20	gat	aat	cca	gtt	tac	tgt	ggt	tat	atc	aaa	ttt	aag	gac	agc	cta	ttt	720
	Asp	Asn	Pro	Val	Tyr	Cys	Gly	Tyr	Ile	Lys	Phe	Lys	Asp	Ser	Leu	Phe	
		225				230					235					240	
25	gaa	ggt	atg	cac	aaa	cca	att	atc	cct	tat	gag	act	tat	tta	aaa	gtt	768
	Glu	Gly	Met	His	Lys	Pro	Ile	Ile	Pro	Tyr	Glu	Thr	Tyr	Leu	Lys	Val	
					245					250					255		
30	caa	aaa	gag	cta	gaa	gaa	aga	caa	cag	cag	act	tat	gaa	aga	aat	aac	816
	Gln	Lys	Glu	Leu	Glu	Glu	Arg	Gln	Gln	Gln	Thr	Tyr	Glu	Arg	Asn	Asn	
				260					265					270			
35	aac	cct	aga	cct	ttc	caa	gct	aaa	tat	atg	ctg	tca	ggg	atg	gca	agg	864
	Asn	Pro	Arg	Pro	Phe	Gln	Ala	Lys	Tyr	Met	Leu	Ser	Gly	Met	Ala	Arg	
			275					280					285				
40	tgc	ggt	tac	tgt	gga	gca	cct	tta	aaa	att	gtt	ctt	ggc	cac	aaa	aga	912
	Cys	Gly	Tyr	Cys	Gly	Ala	Pro	Leu	Lys	Ile	Val	Leu	Gly	His	Lys	Arg	
		290					295					300					
45	aaa	gat	gga	agc	cgc	act	atg	aaa	tat	cac	tgt	gca	aat	aga	ttt	cct	960
	Lys	Asp	Gly	Ser	Arg	Thr	Met	Lys	Tyr	His	Cys	Ala	Asn	Arg	Phe	Pro	
		305				310					315					320	
50	cga	aaa	aca	aaa	gga	att	aca	gta	tat	aat	gac	aat	aaa	aag	tgt	gat	1008
	Arg	Lys	Thr	Lys	Gly	Ile	Thr	Val	Tyr	Asn	Asp	Asn	Lys	Lys	Cys	Asp	
					325					330					335		
55	tca	gga	act	tat	gat	tta	agt	aat	tta	gaa	aat	act	gtt	att	gac	aac	1056
	Ser	Gly	Thr	Tyr	Asp	Leu	Ser	Asn	Leu	Glu	Asn	Thr	Val	Ile	Asp	Asn	
				340					345					350			
60	ctg	att	gga	ttt	caa	gaa	aat	aat	gac	tcc	tta	ttg	aaa	att	atc	aat	1104
	Leu	Ile	Gly	Phe	Gln	Glu	Asn	Asn	Asp	Ser	Leu	Leu	Lys	Ile	Ile	Asn	
			355					360					365				
65	ggc	aac	aac	caa	cct	att	ctt	gat	act	tgc	tca	ttt	aaa	aag	caa	att	1152
	Gly	Asn	Asn	Gln	Pro	Ile	Leu	Asp	Thr	Ser	Ser	Phe	Lys	Lys	Gln	Ile	
		370					375					380					
70	tca	cag	atc	gat	aaa	aaa	ata	caa	aag	aac	tct	gat	ttg	tac	cta	aat	1200
	Ser	Gln	Ile	Asp	Lys	Lys	Ile	Gln	Lys	Asn	Ser	Asp	Leu	Tyr	Leu	Asn	
		385				390					395					400	
75	gat	ttt	atc	act	atg	gat	gag	ttg	aaa	gat	cgt	act	gat	tcc	ctt	cag	1248
	Asp	Phe	Ile	Thr	Met	Asp	Glu	Leu	Lys	Asp	Arg	Thr	Asp	Ser	Leu	Gln	
					405					410					415		
80	gct	gag	aaa	aag	ctg	ctt	aaa	gct	aag	att	agc	gaa	aat	aaa	ttt	aat	1296
	Ala	Glu	Lys	Lys	Leu	Leu	Lys	Ala	Lys	Ile	Ser	Glu	Asn	Lys	Phe	Asn	
				420				425						430			
85	gac	tct	act	gat	gtt	ttt	gag	tta	gtt	aaa	act	cag	ttg	ggc	tca	att	1344

93

	Asp	Ser	Thr	Asp	Val	Phe	Glu	Leu	Val	Lys	Thr	Gln	Leu	Gly	Ser	Ile	
			435					440					445				
5	ccg	att	aat	gaa	cta	tca	tat	gat	aat	aaa	aag	aaa	atc	gtc	aac	aac	1392
	Pro	Ile	Asn	Glu	Leu	Ser	Tyr	Asp	Asn	Lys	Lys	Lys	Ile	Val	Asn	Asn	
		450					455					460					
10	ctt	gta	tca	aag	gtt	gat	gtt	act	gct	gat	aat	gta	gat	atc	ata	ttt	1440
	Leu	Val	Ser	Lys	Val	Asp	Val	Thr	Ala	Asp	Asn	Val	Asp	Ile	Ile	Phe	
	465					470					475					480	
	aaa	ttc	caa	ctc	gct	taa											1458
	Lys	Phe	Gln	Leu	Ala												
					485												
15																	
	<210>	108															
	<211>	485															
	<212>	PRT															
20	<213>	Bacteriophage TP901-1															
	<400>	108															
25	Met	Thr	Lys	Lys	Val	Ala	Ile	Tyr	Thr	Arg	Val	Ser	Thr	Thr	Asn	Gln	
	1				5					10					15		
	Ala	Glu	Glu	Gly	Phe	Ser	Ile	Asp	Glu	Gln	Ile	Asp	Arg	Leu	Thr	Lys	
				20					25					30			
30	Tyr	Ala	Glu	Ala	Met	Gly	Trp	Gln	Val	Ser	Asp	Thr	Tyr	Thr	Asp	Ala	
		35						40					45				
	Gly	Phe	Ser	Gly	Ala	Lys	Leu	Glu	Arg	Pro	Ala	Met	Gln	Arg	Leu	Ile	
		50					55					60					
35	Asn	Asp	Ile	Glu	Asn	Lys	Ala	Phe	Asp	Thr	Val	Leu	Val	Tyr	Lys	Leu	
	65					70					75					80	
	Asp	Arg	Leu	Ser	Arg	Ser	Val	Arg	Asp	Thr	Leu	Tyr	Leu	Val	Lys	Asp	
					85					90					95		
40	Val	Phe	Thr	Lys	Asn	Lys	Ile	Asp	Phe	Ile	Ser	Leu	Asn	Glu	Ser	Ile	
				100					105					110			
45	Asp	Thr	Ser	Ser	Ala	Met	Gly	Ser	Leu	Phe	Leu	Thr	Ile	Leu	Ser	Ala	
		115						120					125				
	Ile	Asn	Glu	Phe	Glu	Arg	Glu	Asn	Ile	Lys	Glu	Arg	Met	Thr	Met	Gly	
		130					135					140					
50	Lys	Leu	Gly	Arg	Ala	Lys	Ser	Gly	Lys	Ser	Met	Met	Trp	Thr	Lys	Thr	
	145					150					155					160	
	Ala	Phe	Gly	Tyr	Tyr	His	Asn	Arg	Lys	Thr	Gly	Ile	Leu	Glu	Ile	Val	
					165					170					175		
55	Pro	Leu	Gln	Ala	Thr	Ile	Val	Glu	Gln	Ile	Phe	Thr	Asp	Tyr	Leu	Ser	
				180					185					190			
60	Gly	Ile	Ser	Leu	Thr	Lys	Leu	Arg	Asp	Lys	Leu	Asn	Glu	Ser	Gly	His	
		195						200					205				
	Ile	Gly	Lys	Asp	Ile	Pro	Trp	Ser	Tyr	Arg	Thr	Leu	Arg	Gln	Thr	Leu	
		210					215					220					
65	Asp	Asn	Pro	Val	Tyr	Cys	Gly	Tyr	Ile	Lys	Phe	Lys	Asp	Ser	Leu	Phe	
	225					230					235					240	

94

	Glu	Gly	Met	His	Lys	Pro	Ile	Ile	Pro	Tyr	Glu	Thr	Tyr	Leu	Lys	Val
					245					250					255	
5	Gln	Lys	Glu	Leu	Glu	Glu	Arg	Gln	Gln	Gln	Thr	Tyr	Glu	Arg	Asn	Asn
				260				265						270		
	Asn	Pro	Arg	Pro	Phe	Gln	Ala	Lys	Tyr	Met	Leu	Ser	Gly	Met	Ala	Arg
			275					280					285			
10	Cys	Gly	Tyr	Cys	Gly	Ala	Pro	Leu	Lys	Ile	Val	Leu	Gly	His	Lys	Arg
		290					295					300				
	Lys	Asp	Gly	Ser	Arg	Thr	Met	Lys	Tyr	His	Cys	Ala	Asn	Arg	Phe	Pro
15	305					310					315					320
	Arg	Lys	Thr	Lys	Gly	Ile	Thr	Val	Tyr	Asn	Asp	Asn	Lys	Lys	Cys	Asp
					325					330					335	
20	Ser	Gly	Thr	Tyr	Asp	Leu	Ser	Asn	Leu	Glu	Asn	Thr	Val	Ile	Asp	Asn
				340					345					350		
	Leu	Ile	Gly	Phe	Gln	Glu	Asn	Asn	Asp	Ser	Leu	Leu	Lys	Ile	Ile	Asn
			355					360					365			
25	Gly	Asn	Asn	Gln	Pro	Ile	Leu	Asp	Thr	Ser	Ser	Phe	Lys	Lys	Gln	Ile
		370					375					380				
	Ser	Gln	Ile	Asp	Lys	Lys	Ile	Gln	Lys	Asn	Ser	Asp	Leu	Tyr	Leu	Asn
30	385					390					395					400
	Asp	Phe	Ile	Thr	Met	Asp	Glu	Leu	Lys	Asp	Arg	Thr	Asp	Ser	Leu	Gln
					405					410					415	
35	Ala	Glu	Lys	Lys	Leu	Leu	Lys	Ala	Lys	Ile	Ser	Glu	Asn	Lys	Phe	Asn
				420					425					430		
	Asp	Ser	Thr	Asp	Val	Phe	Glu	Leu	Val	Lys	Thr	Gln	Leu	Gly	Ser	Ile
			435					440					445			
40	Pro	Ile	Asn	Glu	Leu	Ser	Tyr	Asp	Asn	Lys	Lys	Lys	Ile	Val	Asn	Asn
		450					455					460				
	Leu	Val	Ser	Lys	Val	Asp	Val	Thr	Ala	Asp	Asn	Val	Asp	Ile	Ile	Phe
45	465					470					475					480
	Lys	Phe	Gln	Leu	Ala											
					485											

Fig. 1

1/9

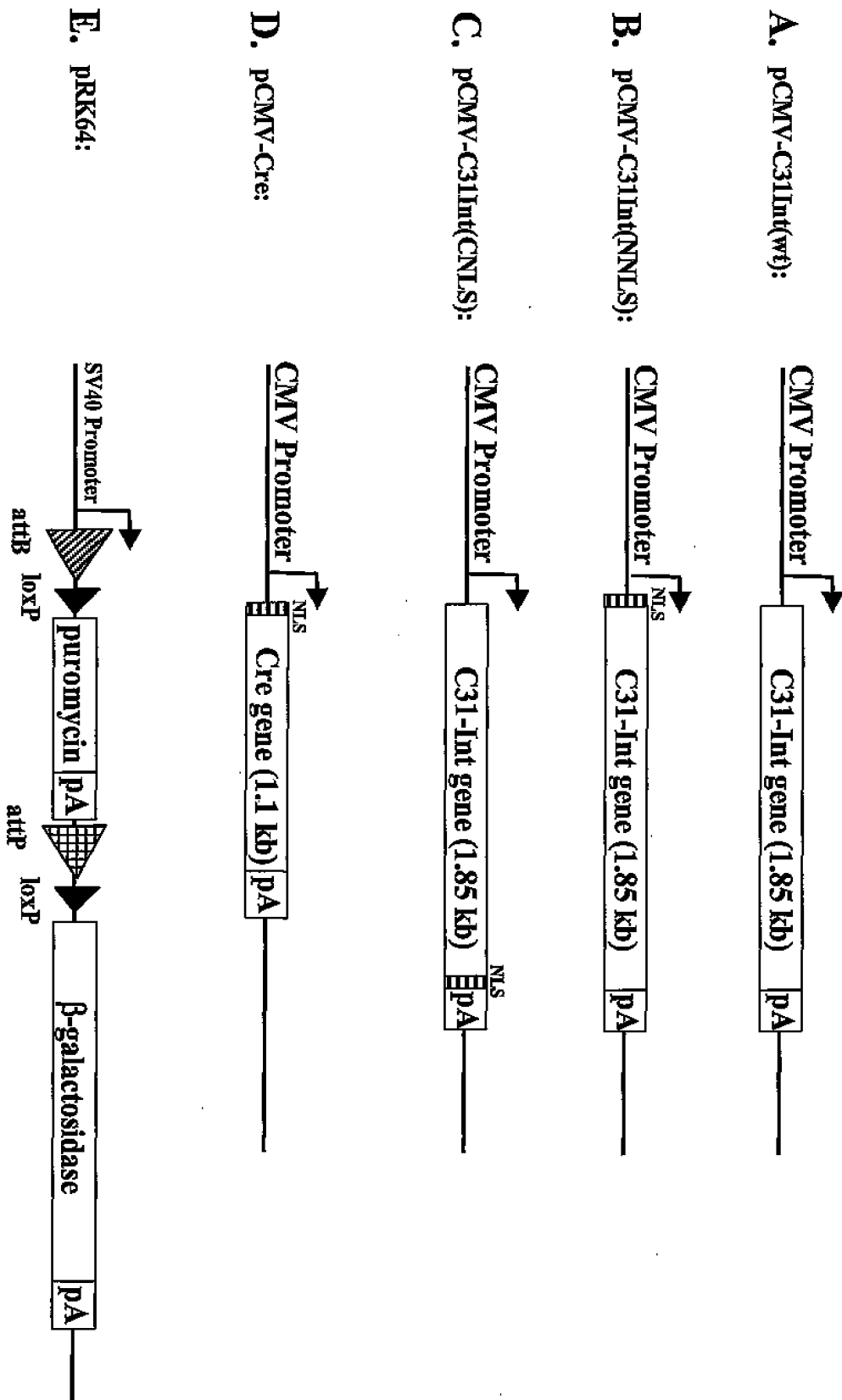


Fig. 2

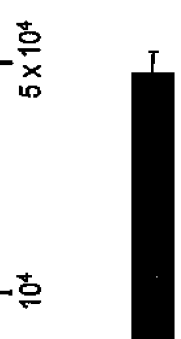

















Sample	RLU (β Gal)	RLU (Luciferase)	RLU $\times 10^5$ (Gal/Luc)	RLU $\times 10^5$ (Gal/Luc)	relative activity
1) pUC19 only	433 \pm 37	1818700 \pm 328970	24 \pm 4		1
2) pRK64(Δ Cre)	784932 \pm 290524	1657062 \pm 526562	46975 \pm 3696		0.004
3) pRK64 only	1936 \pm 262	988144 \pm 175116	204 \pm 62		
4) pCMV-C31Int(wt) 0.5 ng	73318 \pm 19084	677861 \pm 145341	10774 \pm 972		0.23
5) pCMV-C31Int(wt) 1 ng	75838 \pm 12628	527237 \pm 53846	14412 \pm 2050		0.3
6) pCMV-C31Int(NNLS) 0.5 ng	158402 \pm 75870	2560450 \pm 736186	15104 \pm 3041		0.32
7) pCMV-C31Int(NNLS) 1 ng	206857 \pm 76733	2677621 \pm 504285	17029 \pm 2246		0.36
8) pCMV-C31Int(CNLS) 0.5 ng	274192 \pm 78937	1173932 \pm 291315	23299 \pm 3194		0.5
9) pCMV-C31Int(CNLS) 1 ng	262169 \pm 60583	864752 \pm 229935	30560 \pm 1585		0.65
10) pCMV-Cre(NNLS) 0.5 ng	231200 \pm 96741	763121 \pm 280687	29595 \pm 4632		0.63
11) pCMV-Cre(NNLS) 1 ng	297760 \pm 83363	868905 \pm 196404	33872 \pm 2609		0.72

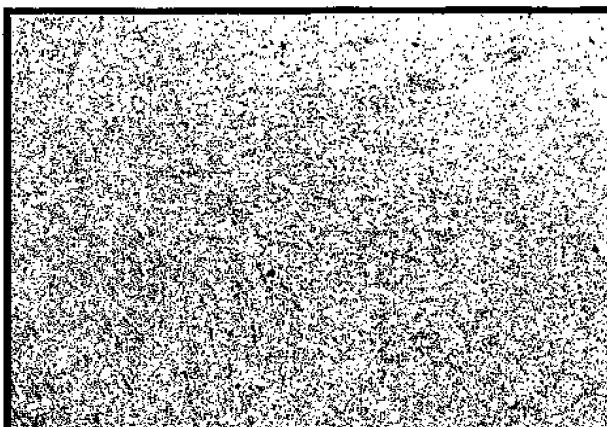
Fig. 3

	RLU (β Gal)	RLU (Luciferase)	RLU $\times 10^5$ (Gal/Luci)
1) pPGKnifD (reporter) only	1324 \pm 876	3631598 \pm 903012	34 \pm 18
2) pCMV-XisA 25 ng	4650 \pm 2273	2741969 \pm 667568	164 \pm 54
3) pCMV-XisA 100 ng	17529 \pm 9304	3798872 \pm 1288020	443 \pm 151
4) pCMV-XisA(NNLS) 25 ng	4060 \pm 1376	2471695 \pm 611351	163 \pm 36
5) pCMV-XisA(NNLS) 100 ng	17801 \pm 3892	3570103 \pm 750628	500 \pm 65
6) pPGKattA (reporter) only	754 \pm 70	195822 \pm 81858	755 \pm 601
7) pCMV-SSV 10 ng	925 \pm 273	119043 \pm 67451	906 \pm 316
8) pCMV-SSV 20 ng	1033 \pm 270	122557 \pm 30054	879 \pm 291
9) pCMV-SSV(NNLS) 10 ng	1108 \pm 367	174380 \pm 58876	694 \pm 345
10) pCMV-SSV(NNLS) 20 ng	1306 \pm 383	211182 \pm 101011	874 \pm 741

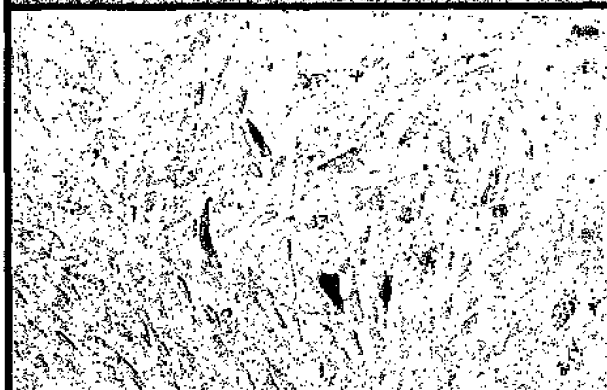
Fig. 4

Sample	RLU (βGal)	RLU (βGal)	relative activity
1) pUC19 only	754 ± 44		
2) pCMV-C31Int(wt) 32 ng	1386 ± 174		0.04
3) pCMV-C31Int(wt) 64 ng	3783 ± 1537		0.1
4) pCMV-C31Int(NNLS) 32 ng	7125 ± 1474		0.19
5) pCMV-C31Int(NNLS) 64 ng	8206 ± 2210		0.22
6) pCMV-C31Int(CNLS) 32 ng	17624 ± 5578		0.48
7) pCMV-C31Int(CNLS) 64 ng	28849 ± 6623		0.78
8) pCMV-Cre(NNLS) 32 ng	27064 ± 3769		0.73
9) pCMV-Cre(NNLS) 64 ng	36823 ± 3993		1

**A.: Nontransfected
control**



B.: pCMV-Cre



C.: pCMV-C31Int(NLS)

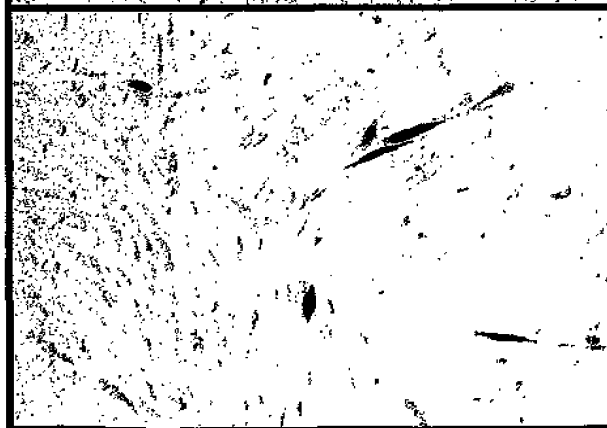


Fig. 5

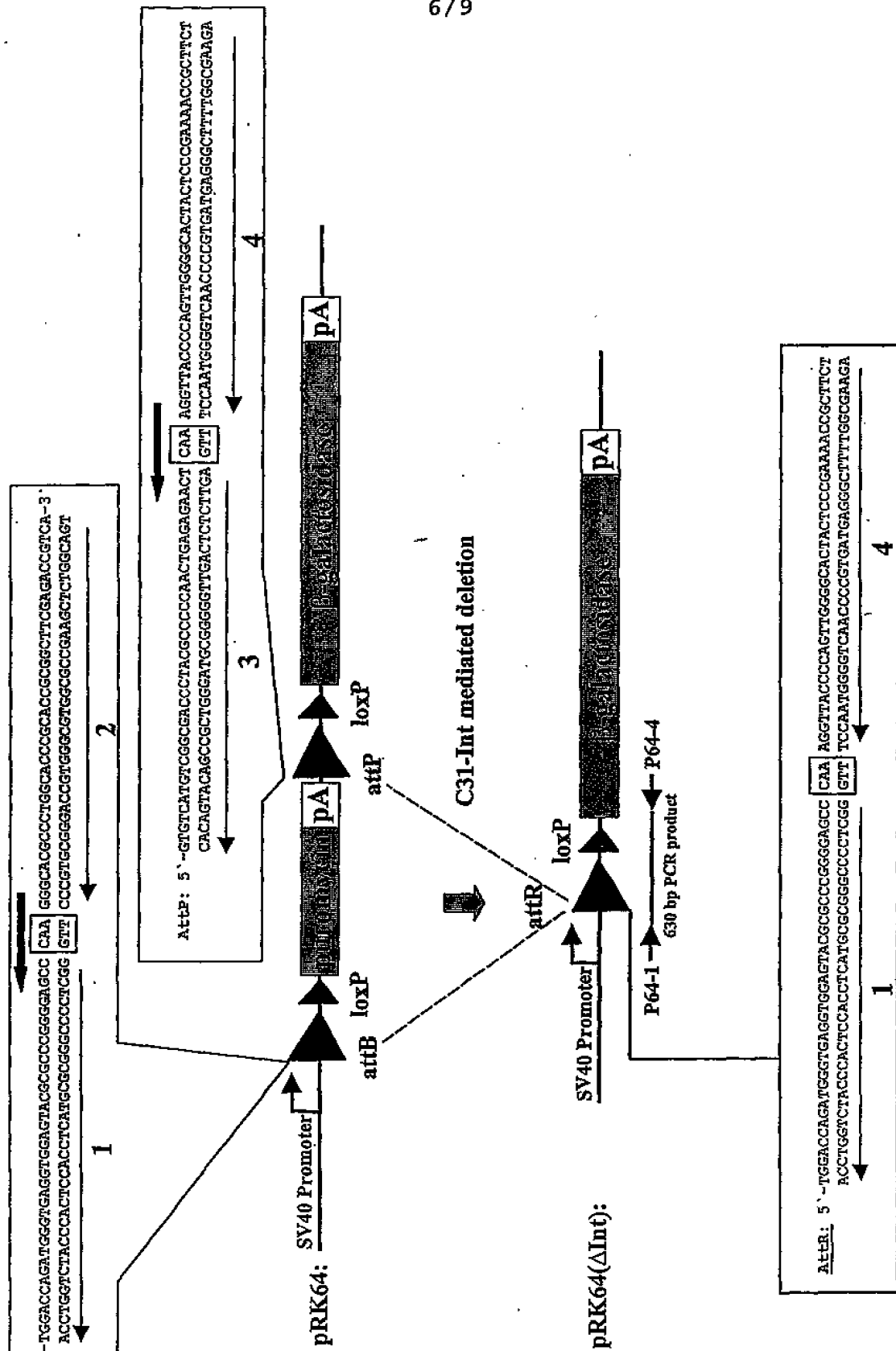


Fig. 6

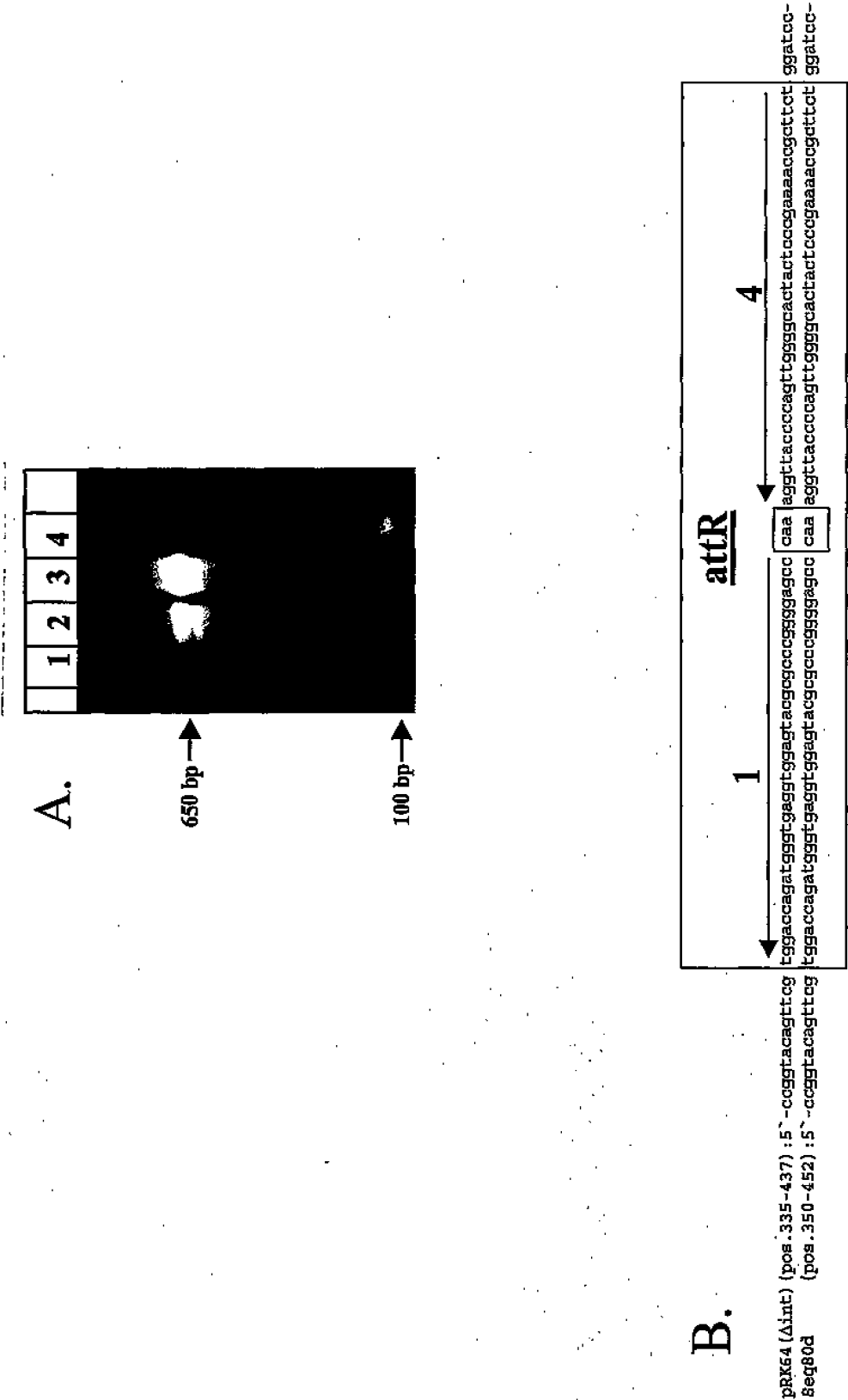
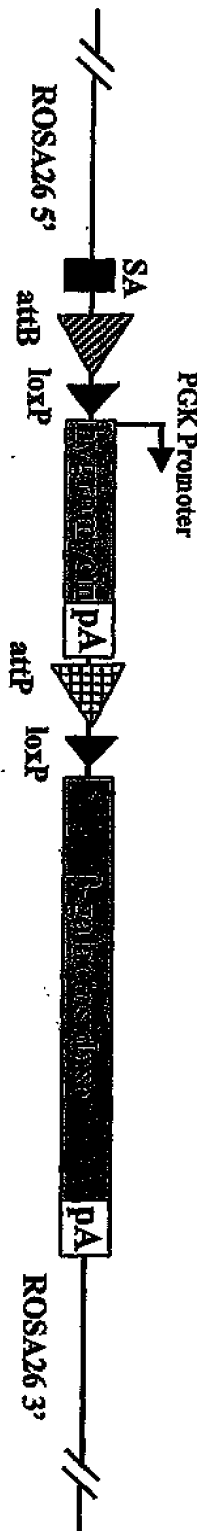


Fig. 7

Fig. 8



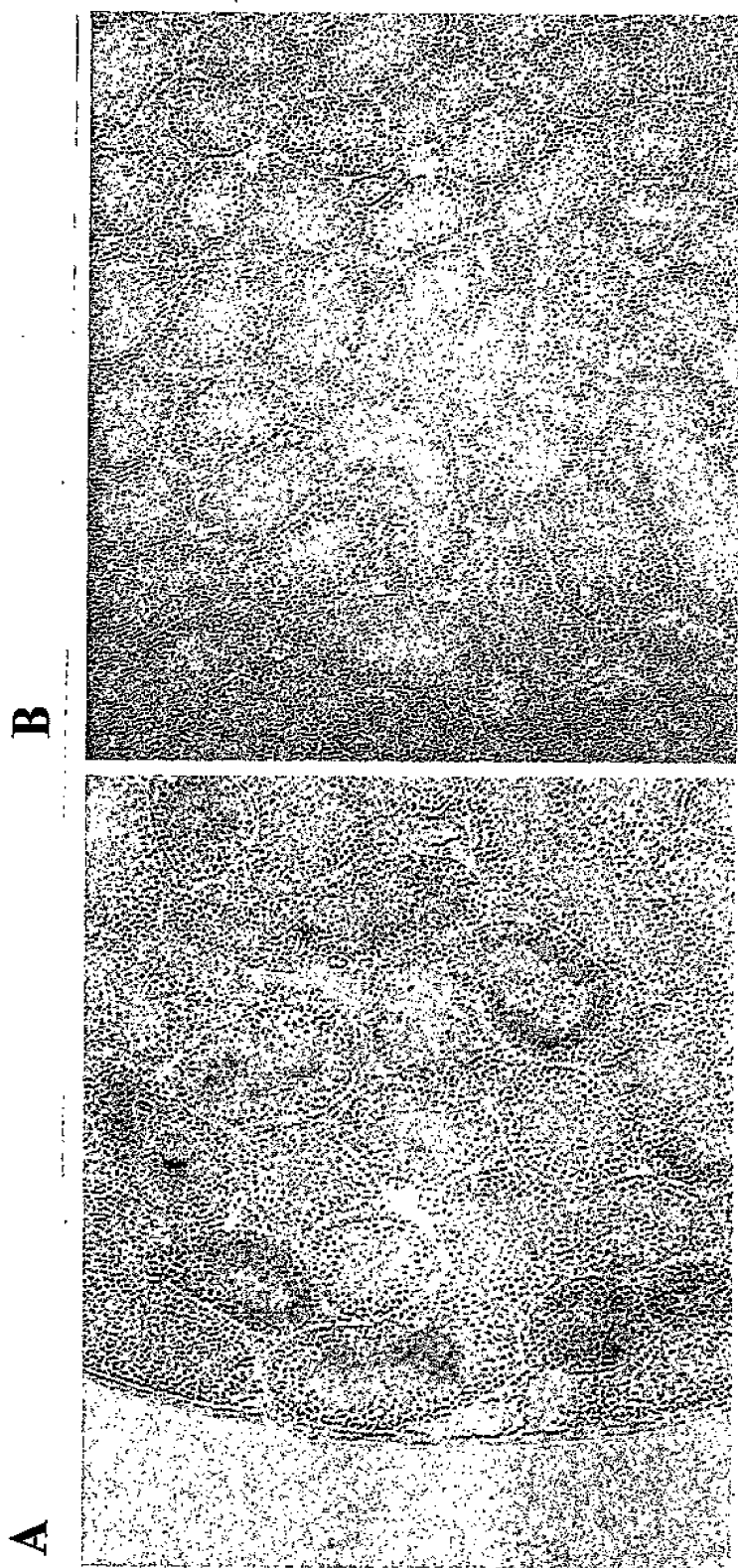


Fig. 9